

From
DEPARTMENT OF MOLECULAR MEDICINE
AND SURGERY
Karolinska Institutet, Stockholm, Sweden

**COMPUTED TOMOGRAPHY
BASED ASSESSMENT OF
TREATMENT RESPONSE IN
SOLID TUMORS**

Chikako Suzuki



**Karolinska
Institutet**

Stockholm 2012

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet. Printed by Larserics

© Chikako Suzuki, 2012

ISBN 978-91-7457-812-6

*For an idea that does not first seem insane,
there is no hope.*

Albert Einstein

It always seems impossible until it's done.

Nelson Mandela

To my mother Tomoko and my daughter Sakiko

ABSTRACT

A “substantial evidence” of effectiveness is required for new cancer treatment regimens to be approved. Objective tumor shrinkage/enlargement has been adopted as an indicator of drugs efficacy. The change of tumor size is assessed and quantified by various radiological techniques; most commonly computed tomography (CT). A high accuracy and reproducibility is, for obvious reasons, necessary in order to achieve a meaningful evaluation of such studies. For that purpose, the World Health Organization criteria (WHO-criteria) were launched in 1979 followed by the Response Evaluation Criteria In Solid Tumors (RECIST) in 2000 and the updated version RECIST 1.1 in 2009. There are, however, still several steps that may deteriorate consistencies.

The purpose of this thesis was to investigate causes that may affect inconsistency in evaluation procedure according to RECIST (study I and II) and to explore the percentage of tumor size change at the first follow-up CT as the potential new surrogate indicators for OS in patient with metastatic colorectal cancer (mCRC) (study III) and in patient with metastatic breast cancer (MBC) (study IV).

The number of discordant cases increased gradually when assessing fewer target lesions. Measuring fewer than four target lesions might cause discrepancies when more than five target lesions were present (study I). Interobserver variation using RECIST and WHO-criteria were moderate: 0.53 (95%CI 0.33 - 0.72) and 0.60 (0.39 – 0.80), respectively. Intraobserver variation using RECIST and WHO-criteria were substantial to perfect that ranged between 0.76 – 0.96 and 0.86 – 0.91, respectively (study II).

The initial change in tumor size 8 weeks after initiation of chemotherapy was prognostic for PFS: Hazard Ratio (HR) 2.21, 95%CI 1.97 – 2.49, and OS: HR2.01, 95%CI 1.75 – 2.31, in mCRC (study III). The initial change in tumor size also correlated with OS in MBC (study IV). A marked difference in OS between patients with or without new lesion was demonstrated in mCRC: HR 3.77, 95%CI 2.08 – 6.83 (study III) and in MBC: HR 4.29, 95%CI 2.44 – 7.53 (study IV).

In conclusion, the current tumor response evaluation criteria are associated with several subjective steps that may cause inconsistent results. The initial change in tumor size at the first follow-up CT may provide an alternative surrogate outcome. The findings obtained in this thesis may improve the development of future response evaluation criteria.

LIST OF PUBLICATIONS

- I. DARKEH, M. H., SUZUKI, C. & TORKZAD, M. R. 2009.
The minimum number of target lesions that need to be measured to be representative of the total number of target lesions (according to RECIST).
Br J Radiol, 82, 681-6.
- II. SUZUKI, C., TORKZAD, M. R., JACOBSSON, H., ASTROM, G., SUNDIN, A., HATSCHEK, T., FUJII, H. & BLOMQVIST, L. 2010.
Interobserver and intraobserver variability in the response evaluation of cancer therapy according to RECIST and WHO-criteria.
Acta Oncol, 49, 509-14.
- III. SUZUKI, C., BLOMQVIST, L., SUNDIN, A., JACOBSSON, H., BYSTROM, P., BERGLUND, A., NYGREN, P. & GLIMELIUS, B. 2012.
The initial change in tumor size predicts response and survival in patients with metastatic colorectal cancer treated with combination chemotherapy.
Ann Oncol, 23, 948-54.
- IV. SUZUKI, C., BLOMQVIST, L., HATSCHEK, T., CARLSSON, L., EINBEIGI, Z., LINDERHOLM, B., LINDH, B., LOMAN, N., MALMBERG, M., ROTSTEIN, S., SÖDERBERG, M., SUNDQUIST, M., WALZ, T., ÅSTRÖM, G., FUJII, H., JACOBSSON, H., GLIMELIUS, B.
Impact of early response on overall survival in metastatic breast cancer patients treated with combination chemotherapy within a randomized phase III trial
(submitted manuscript)
- V. *Appendix*
SUZUKI, C., JACOBSSON, H., HATSCHEK, T., TORKZAD, M. R., BODEN, K., ERIKSSON-ALM, Y., BERG, E., FUJII, H., KUBO, A. & BLOMQVIST, L. 2008.
Radiologic measurements of tumor response to treatment: practical approaches and limitations.
Radiographics, 28, 329-44. *(Review Article)*

CONTENTS

1	INTRODUCTION.....	1
1.1	Epidemiology and Economy of Cancer	2
1.2	Evaluation of effectiveness of cancer therapy	3
1.2.1	Clinical Trials	3
1.2.2	Biomarkers in Oncology	4
1.2.3	Survival as an Endpoint and it's Surrogate Endpoints	5
1.3	Radiological Imaging Based Response Evaluation	6
1.3.1	WHO-criteria and RECIST	6
1.3.2	Background of Threshold Setting	9
1.3.3	Background of Target Lesion Number in RECIST (1.1)....	10
1.3.4	Revised RECIST guideline version 1.1 (RECIST 1.1).....	11
1.4	Tumor measurements beyond WHO-criteria and RECIST	13
1.4.1	Waterfall analyses	13
1.4.2	Functional Imaging Based Evaluation	13
1.5	Emerging Image technology and cancer drugs.....	15
2	AIMS	17
2.1	Study I.....	17
2.2	Study II.....	17
2.3	Study III	17
2.4	Study IV	17
3	MATERIALS AND METHODS	19
3.1	Study Subjects.....	19
3.2	Definition of OR according TO RECIST and WHO-criteria.....	19
3.3	Study I.....	19
3.4	Study II.....	20
3.5	Study III and IV	21
3.5.1	Definition of first change	21
3.5.2	Survival analyses	21
3.5.3	Statistical computations.....	24
3.5.4	Patients in Study III.....	25
3.5.5	Patients in Study IV.....	27
4	RESULTS	29
4.1	Study I.....	29
4.2	Study II.....	30
4.3	Study III	32
4.3.1	Relationship between 1 st change and OS in mCRC	32
4.3.2	Relationship between 1 st change and PFS in mCRC	35
4.3.3	The 2nd follow-up examination.....	37
4.4	Study IV	39
4.4.1	The extent of 1 st change in MBC	39
4.4.2	Relationship between 1 st change and OS in MBC	40
4.4.3	The Cox regression analysis of OS in MBC	42
5	DISCUSSION.....	43
5.1	To measure or not to measure, that is the problem.....	43
5.2	Clinical implications, What matters most?.....	44

5.3	Limitations	46
5.4	Future Perspectives.....	48
6	Conclusions	49
7	Appendix	50
7.1	Illustrative chart for tumor evaluation	50
8	Acknowledgements	53
9	References	55

LIST OF ABBREVIATIONS

AA	accelerated approval
AIDS	acquired immune deficiency syndrome
BOR	best overall response
CAD	computer aided detection
CR	complete response
CRC	colorectal cancer
CRF	case record form
CT	computed tomography
DWI	diffusion weighted image
EGFR	the epidermal growth factor receptor
EORTC	The European Organization for Research and Treatment of Cancer
FDA	The U.S. Food and Drug Administration
¹⁸ F-FDG	[¹⁸ F]- 2-fluoro-2-deoxy-D-glucose
HER-2/neu	human epidermal growth factor 2
HR	hazard ratio
ICC	the intra-class correlation coefficient
KRAS	Kirsten rat sarcoma viral oncogene homolog
LD	the largest diameter
MBC	metastatic breast cancer
MRI	magnetic resonance imaging
mCRC	metastatic colorectal cancer
NCI	National Cancer Institute of United States
NCI Canada	National Cancer Institute of Canada
ORR	objective response rate
OS	overall survival
PD	progressive disease
PET	positron emission tomography
PFS	progression-free survival
PR	partial response
RECIST	Response Evaluation Criteria In Solid Tumors
ROI	regions of interest

RR	response rate
SD	stable disease
SUV	standardized uptake value
TTP	time to progression
WHO	World Health Organization

1 INTRODUCTION

Clinical trials are mandatory in the evaluation of new tumor treatments. A commonly studied indicator of the effect of an instituted therapy is the change of size of the malignant lesion(s)¹. Both tumor shrinkage (response) and growth (progress) are assumed to provide useful endpoints in clinical trials. The change of size of tumor burden before and after treatment is often assessed and quantified by various techniques; most commonly computed tomography (CT). A high accuracy and reproducibility are, for obvious reasons, necessary in order to achieve a meaningful evaluation of such studies.

In 1940's to 1950's, the radiological techniques were still primitive. Tumor size was mainly measured by oncologists placing a ruler or caliper over a deep-seated irregular mass through the muscle, fat and skin². There was, obviously, an inevitable factor of human error in these measurements which could cause serious effects on the conclusion of a clinical trial. A simulation experiment in which balls of different sizes were measured through a layer of foam rubber demonstrated that tumor measurement error increased with the size of tumor and that variation between investigators was larger than the one within investigator. On the basis of simulation results, a 50% reduction in the product of perpendicular diameters of tumor is employed to keep an objective response rate (ORR) of 5 to 10% due to human error in tumor measurement². In 1960's to 70's, it became more common to obtain tumor size on "roentgenograms", such as chest and abdominal X-ray². In order to reduce the confusion caused by various methods used *ad hoc* for therapy evaluation of solid tumors with radiological methods, the World Health Organization criteria (WHO-criteria) were launched in 1979^{3,4}. As an extension and modification of these definitions, the Response Evaluation Criteria In Solid Tumors (RECIST) were published in 2000⁵. RECIST has become dominating⁶; an updated version was published as RECIST 1.1 in 2009⁷. By using these criteria, the therapy response is supposed to be assessed in a common way by measuring tumor size before, during and after the treatment. The effectiveness of cancer treatment is then categorized into four statuses: complete response (CR), partial response (PR), stable disease (SD) and progressive disease (PD). The first two categories are combined to define responders and the last two combined to define non-responders¹.

Several studies have shown inconsistencies in different kinds of image based tumor measurements; either RECIST (uni-dimensional), WHO-criteria (bi-

dimensional), volume (three-dimensional) or standardized uptake value (SUV) obtained by positron emission tomography (PET), and discussed appropriate measurements for evaluation⁸⁻¹⁵. Few have questioned other factors than the measurement itself causing inconsistent results of treatment response in clinical trials. Neither the stipulation of how many target lesions should be selected according to RECIST nor the threshold settings for categorization of response is supported by any scientific evidence^{2, 16}. Few studies have evaluated the fewest number of target lesions being necessary for an adequate evaluation or the inter- and intraobserver variability when a reader is confronted with a series of radiological images having to select and measure target lesions from various organs^{17, 18}. Several studies have questioned the ORR as a surrogate endpoint for overall survival (OS) since SD might also be due to the treatment effect^{19, 20}, while few has questioned how and why certain thresholds in percent change in size of diameter of tumors have been defined for response classification^{19, 21-24}. Instead, the value of emerging functional imaging techniques, such as PET-CT, has been increasingly discussed in terms of a potential evaluation tool beyond tumor size based on anatomical imaging²⁵⁻²⁷.

The purpose of this thesis was to determine the minimum number of target lesions to be representative of full assessment of target lesions according to RECIST (study I), to examine the intra- and interobserver agreement using two different using two different response evaluation criteria (study II), to determine whether the change in tumor diameters at the first follow-up CT (first change) correlates with outcome in patients with metastatic colorectal cancer (mCRC) (study III) and in patients with metastatic breast cancer (MBC) (study IV).

1.1 EPIDEMIOLOGY AND ECONOMY OF CANCER

Cancer is now the leading cause of death in the world. According to the IARC report, the global burden of cancers is supposed double by 2020 and nearly triple by 2030²⁸. As many as 26 million new cancer diagnoses and up to 17 million deaths worldwide are supposed for 2030 compared with an estimated 12 million new cancer diagnoses and more than 7 million deaths in 2008²⁸. Almost two-thirds of deaths might occur in low- and middle-income countries²⁹⁻³². Many interventions for early detection and treatment might remain inaccessible, thus making cancer a leading cause of mortality for many people in developing countries³³.

Better systemic therapy has contributed to improve prognosis in many cancers. Among patients with metastatic colorectal cancer, the median duration of survival was eight

months without chemotherapy³⁴. With fluorouracil, it increased to 12 months³⁴. The combination of fluorouracil, irinotecan and oxaliplatin extended the median survival to 21 months³⁴. Bevacizumab and cetuximab are likely to prolong median survival beyond 21 months³⁴. A concern when dealing with new systemic therapies is the cost of treatment. Estimated drug costs for a patient with colorectal cancer who receives chemotherapy for eight weeks are about \$100 for fluorouracil, \$10,000 for regimens containing irinotecan or oxaliplatin and \$30,000 for regimens containing bevacizumab or cetuximab³⁴. In the United States approximately 56,000 patients need chemotherapy for colorectal cancer in a year, and the drug costs for an eight-week course of treatment will be approximately \$666 million or \$1.2 billion with the addition of monoclonal-antibody therapy³⁴. Obviously, these astronomical costs are a threat to the public health care systems. But why have the costs become so high? The cost for drug development is an important reason. In the United States the average cost of bringing a drug to the market increased from \$318 million in 1991 to \$802 million in 2003³⁵. Another possible answer is a lengthy process from initial laboratory and animal testing to phase I, II, III studies for drug approval which may take years³⁶.

1.2 EVALUATION OF EFFECTIVENESS OF CANCER THERAPY

1.2.1 Clinical Trials

Clinical trials are conducted to allow safety and efficacy data to be collected for various medical interventions; e.g. drugs, diagnostic methods and devices, treatment regimens. Clinical trials are mandatory for evaluation and implementation of new cancer treatments^{37,38}. The whole procedure mainly consists of three to four phases. Each phase have a different purpose³⁹.

- Phase I: Screening for safety. To determine side effects and an appropriate dose for later study. Conducted in a small group of people (<100)
- Phase II: Efficacy assessment in people with a particular condition. Establish the testing protocol in phase III trials^{20, 40-42}. Primary endpoint: objective response rate (ORR). Conducted in a larger group of people (100-300)
- Phase III: Studies focuses on comparison of a new, experimental treatment with an existing standard treatment. To confirm findings of phase III trials⁴¹. Primary endpoint: TTP/PFS, OS. Conducted in large groups of people (500-3000)

Phase IV: “Post-approval” studies to delineate additional information including the drug’s risks, benefits and optimal use.

The average probability that a drug candidate will successfully pass clinical phase I is in the range of 75%; the respective values for phase II and III are 50% and 65%⁴³. In total, the cumulative probability that a leading drug candidate will successfully proceed from the preclinical phase to approval is about 8%⁴⁴. This high failure rate, especially in phase III trials, may be due to inappropriate design of phase III trials based on the result of phase II trials. RR, which is the usual primary endpoint of phase II, has weak correlation with OS thus may indicate neither reliably drug activity nor subsequent phase III trial is warranted⁴⁰.

1.2.2 Biomarkers in Oncology

In oncology, there are three types of biomarkers which are of importance: prognostic, predictive and early response biomarkers^{44,45}. Prognostic biomarkers predict the likely course of disease, irrespective of treatment; for example, lymph node involvement predicts a poor outcome in the management of solid tumors, even though treatment can prolong survival of patients with or without evidence of nodal involvement. These biomarkers are valuable not only for identifying aggressive disease that requires immediate treatment, but for identifying indolent disease that can safely be left alone³². Predictive biomarkers forecast the likely response to treatment; for example, *KRAS* mutation predicts lack of response to treatment with cetuximab and panitumumab in colon cancer⁴⁶, or hormone receptor status predicts response to endocrine therapies in breast cancer⁴⁷. Furthermore, some biomarkers, such as hormone receptor status and HER2/neu status in breast cancer, are both prognostic and predictive.

Early response biomarkers reveal treatment response or failure earlier than it is possible with conventional response assessment methods. Their use can spare patients from prolonged exposure to ineffective treatments and allow alternative therapies to be tried sooner, and also save money. If they can be served as surrogate endpoints, aiming at replacing a clinical endpoint with a faster and more-sensitive evaluation of the effect of experimental treatments, early response biomarkers can also hasten the progress of clinical trials^{32,45}.

1.2.3 Survival as an Endpoint and it's Surrogate Endpoints

Overall survival (OS) is defined as the time from randomization to death from any cause, and has historically been regarded as the gold standard measure of clinical benefit^{37-39, 48, 49}. OS is favored due to its objectivity, clear indication of benefit, ease and reliability measurement. However, determination of OS requires a larger sample size, a prolonged follow-up period, and may be influenced by other therapies used after patient's participating trial has ended; including crossover to the experimental arm³⁹. This has been of increasing concern because significant advances in the systemic treatment have altered the number and variety of available treatments. To avoid the influence by subsequent treatments and to shorten the time period that requires for response evaluation, time-to-event endpoints, such as progression-free survival (PFS) and the related endpoint time to progression (TTP) have been considered of surrogate endpoints for OS³⁷. PFS is defined as the time elapsed between randomization and tumor progression or death from any cause. TTP is defined as the time elapsed between randomization and disease progression.

To improve access to therapeutics for life-threatening diseases, namely acquired immune deficiency syndrome (AIDS), the U.S. Food and Drug Administration (FDA) established the accelerated approval (AA) besides regular approval in 1992. The use of AA was expanded to oncology in 1996^{38, 39}. AA is less stringent compared to the regular approval and it allows the sponsors to begin marketing relevant drugs on the basis of trials that indentify improvements in surrogate outcomes that are reasonably likely to predict clinical benefit, such as the ORR (= the number of complete and partial response) or PFS interval. These surrogate endpoints have been increasingly used as primary endpoints and are now the most frequently used in randomized clinical trials³⁹. However the degree of association between these surrogate endpoints and OS is sometimes inconsistent^{24 19, 21, 23, 50, 51}.

1.3 RADIOLOGICAL IMAGING BASED RESPONSE EVALUATION

1.3.1 WHO-criteria and RECIST

The development of cancer drugs can be traced back to the Manhattan Project (1945). In this project, a number of laboratories developed various kind of weapons of mass destruction such as an atomic bomb and alkylating agents; nitrogen mustard²⁰. Ever since Gilman et al described the possible value of the nitrogen mustards in the treatment of lymphoma, there has been a rapid and continuous increase in the number of investigations of cancer therapy carried out in the world^{52, 53}(Figure 1). This has resulted in an enormous expansion of the cancer literature.

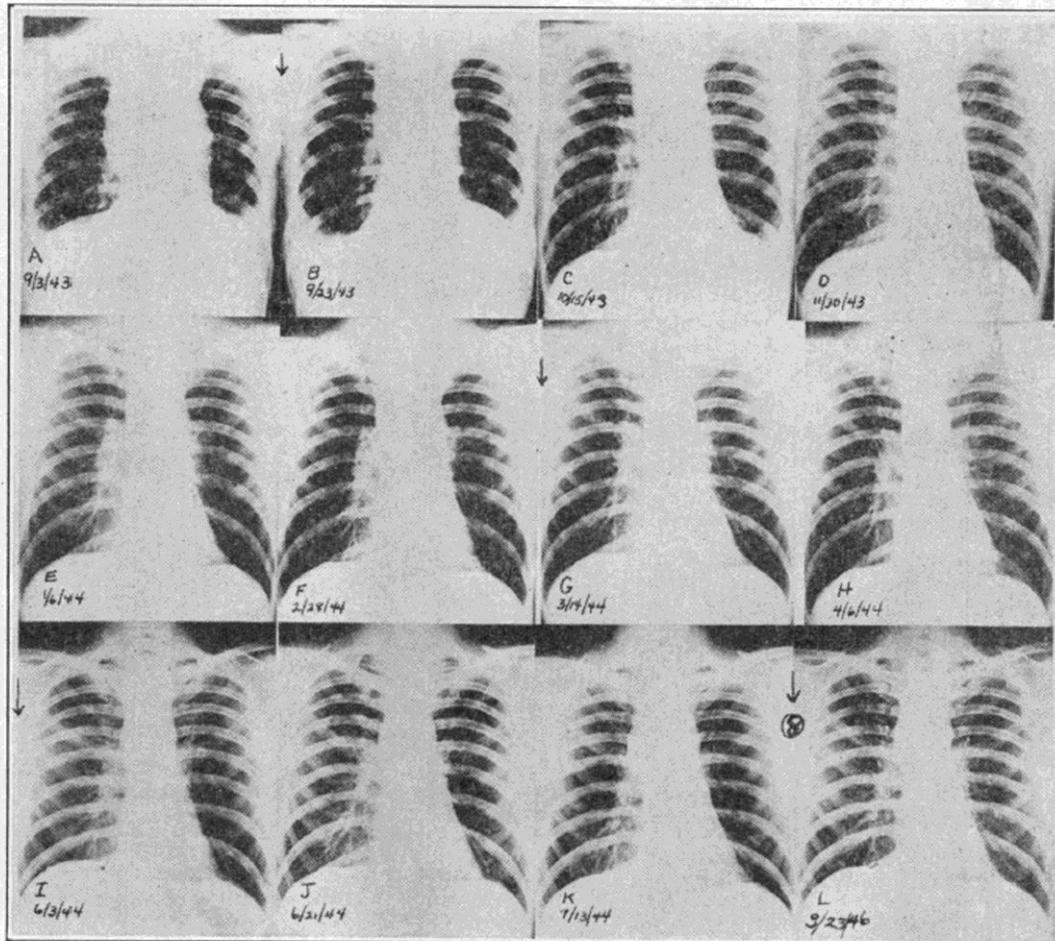


Fig. 2. (case 2).—Roentgenograms showing effect of therapy on the progress of the mediastinal and hilar lesion; A through L represent the pretreatment and subsequent films taken during the period of observation. The arrow inserted between reproductions indicates that a treatment was given; from July 13, 1944 to March 1946 (K and L) eight courses of the drug were administered.

Figure 1. One of the primary example of demonstrating effect of cancer therapy by chest X-rays. A lymphoma patient with mediastinal and hilar lymphadenopathy was treated with nitrogen mustard⁵³. The arrow inserted between chests x-rays indicates that a treatment was given. According to the original article it is described, “sixteen days after the first injection (B) showed a reduction of the mediastinal mass to one-half of its former size”. There was, however, no statement how the mass was measured (reprinted from JAMA 1946⁵³, copyright 2012, by permission of American Medical Association)

However, these investigations were frequently reported in a way which made it difficult for investigators to compare their results with those of others. Therefore it became necessary to develop a “common language” to describe cancer treatment and to agree on internationally acceptable general principles for evaluating data^{3, 4}. In 1979, Miller et al. proposed uniform criteria that standardized the recording and reporting of response, recurrence, and disease-free interval and the grading of acute and sub-acute toxicity in the treatment of solid tumors^{3, 4}. These criteria (WHO-criteria) were based on bi-dimensional measurements (2D, Figure 2) since the tumor measurement was mainly obtained by palpation and it was impossible to measure tumor volume using available imaging technology at that time. These criteria were based on the assumption that the tumor shape is spherical⁸. They have received wide acceptance for reporting the results of cancer treatment.

WHO-criteria do not stipulate the minimum lesion size or the number of lesions to be selected in patients with multiple lesions. Nor do they consider the type of imaging modality that should be used. Application of the WHO criteria includes sources of variability and potential for overestimation⁵⁴⁻⁵⁶.

As a consequence of the limitations and imaging technology developments, the RECIST guidelines were introduced in 2000 by the EORTC, the USNCI, and NCI Canada Clinical Trials Group⁵. The primary goal of RECIST was to try to unify the various modifications of WHO-criteria in order to allow for meaningful comparisons between studies. This included:

- 1) The need to maintain the four categories of responses: CR, PR, SD, and PD;
- 2) The need to maintain the same definition of PR so that favorable results of future therapies can be compared with WHO-criteria even though the measurements will be different; and
- 3) The need to modify the definition of PD.

There are five major differences between the RECIST and WHO-criteria

- 1) Adopting unidimensional (1D) measurements (Figure 2): this encourages measuring more lesions and minimizing labor⁸;
- 2) Stipulating the type of imaging to be used;
- 3) Defining the type of tumors that should or should not be chosen;
- 4) Specifying the number of tumor lesions for assessment; up to 5 lesions in the same organ and total 10 lesions in a patient and

5) Making the cut-off point for definition of PD larger.

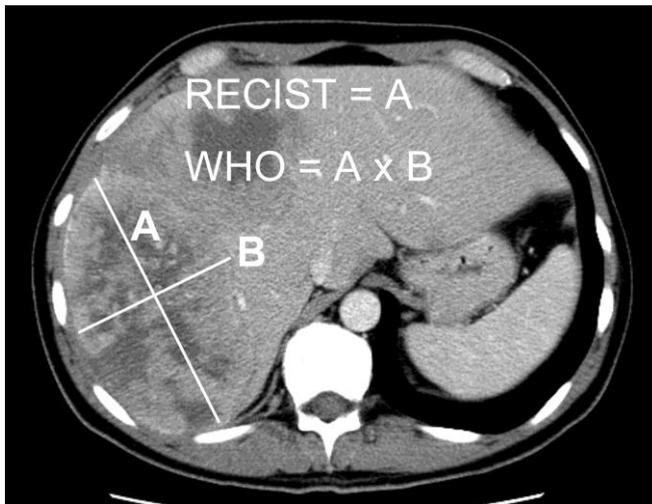


Figure 2. Measuring the size of a liver lesion according to RECIST: the longest diameter (A), and WHO-criteria: the product of A x B (the longest perpendicular diameter).

In spite of these differences, various studies demonstrated concordance between WHO criteria and RECIST for ORR^{8, 57, 58}. However, discrepancies for time to progression could be demonstrated since RECIST requires a larger increase of lesions and longer delay to detect disease progression than WHO-criteria does^{13, 59, 60}.

The evaluation process contains of four main steps: (i) selection of target lesions, (ii) measurement of target lesions, (iii) identification of new lesions or progression of non-target lesions and (iv) categorization of effect, best overall response (BOR) as PD, SD, PR or CR.

The number of patient who reaches CR or PR in study cohort is determined as RR and used for efficacy assessment mainly in phase II studies. PD is used to obtain PFS and TTP.

1.3.2 Background of Threshold Setting

Both tumor shrinkage (response) and tumor increase (disease progression) are useful endpoints in clinical trials¹. Since WHO-criteria were published in 1979, a decrease of >50% and an increase of >25% in the product of the largest and the perpendicular tumor diameters have been used as cut-off values in recording a BOR. There was no biological or radiological background to these original thresholds; instead, it was a way to minimize errors in reporting changes in size of spheres under foam rubber when the measurement of tumor was mainly obtained by palpation². Placing a ruler or caliper over a lump and attempting to estimate its size, there is clearly the inevitable factor of human error. On the basis of measurement of 12 simulated tumor mass models, the 50% reduction criterion was employed to keep the human error in ORR evaluation between 5 and 10 % (Figure 3)².

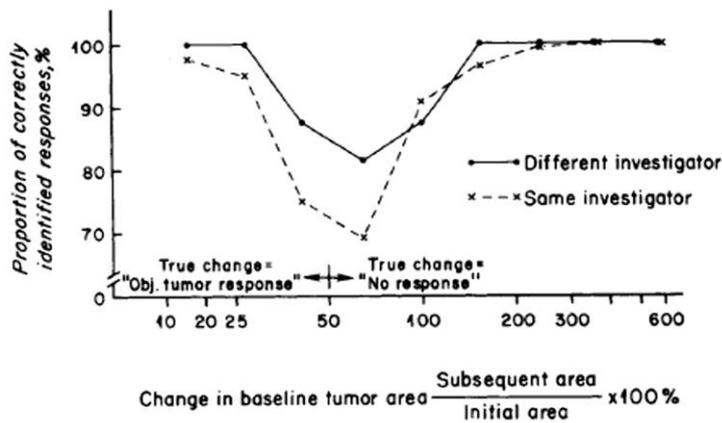


FIG. 2A. Frequency of correctly reported tumor changes (smaller tumors 1-6 only).

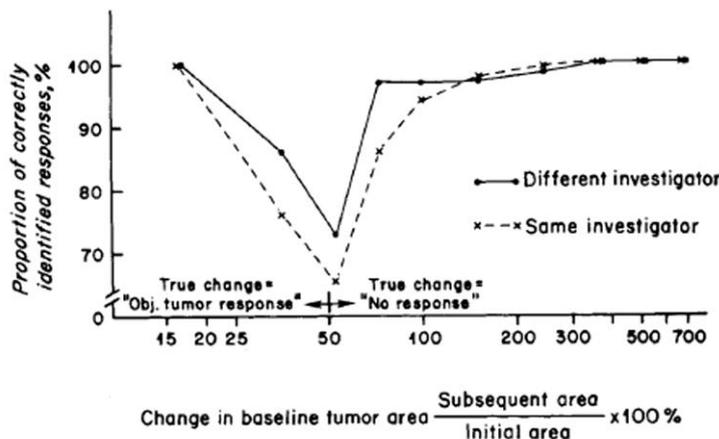


FIG. 2B. Frequency of correctly reported tumor changes (larger tumors 7-12 only).

Figure 3. Frequency of correctly reported tumor changes (reprinted from Cancer 1976 (36), Moertel et al², copyright 2012, by permission of John Wiley and Sons)

When RECIST was determined, these cut-off values were modified after a mathematical conversion from bi- to uni-dimensional measurements in attempt to simplify tumor measurements^{8,61}. Additionally, the cut-off value for progression went from a 40% increase in volume using WHO-criteria to a 73% increase in volume using RECIST^{8,62} (Table 1). It is obvious that imaging techniques have evolved during the last decades and more accurate estimations can be obtained. Nevertheless, the choice of these thresholds is arbitrary without any biological background⁴⁰. However, thresholds for anatomical imaging based response evaluation criteria are difficult to abandon.

Table 1. Equivalent change in diameter, diameters and volume for spherical lesions⁸

Bold letters indicate for change in tumor size to meet response and progress in RECIST and WHO-criteria.

	Change in diameter (1D, RECIST)	Change in diameters (2D, WHO)	Change in volume (3D)
Decrease (response)	30%	50%	65%
Increase (progress)	12%	25%	40%
	20%	44%	73%

Recent studies on the measurement variability of lung lesions on the CT reveal that an average relative difference of 10% in 1-D measurement^{63,64}. Thus a change less than 10% in 1-D can be a result of measurement error alone⁶⁴.

1.3.3 Background of Target Lesion Number in RECIST (1.1)

To solve one problem using WHO-criteria, RECIST stipulated the number of lesions to be recorded: all measurable lesions up to a maximum of five lesions per organ and 10 lesions in total⁵. Again, there is no biological background rather than an effort to minimize time consumption as well as labor in the clinical trial^{35,65,66}. The number of lesions measured on response assessment was revised and reduced in to a maximum of two lesions per organ and 5 lesions in total in RECIST 1.1 after a simulation study⁶⁷.

It is, however, questionable if a few selected target lesions reflect the entire tumor burden. It is possible that the effect of the treatment will not be identical in all sites of metastases. It is sometimes experienced that during treatment some lesions shrink while other lesions grow. The variability in tumor response measurements is

substantially reduced as increasing numbers of lesions are measured¹⁸. However, resources do not permit radiologists to evaluate every lesion⁶⁷.

1.3.4 Revised RECIST guideline version 1.1 (RECIST 1.1)

Since the RECIST was published in 2000, a number of questions and issues raised and led to the development of a revised version of RECIST, so called RECIST 1.1⁷. Amongst these questions and issues were whether less than 10 lesions can be assessed without affecting the overall assigned response for patients or conclusion about activity; how to apply RECIST in randomized phase III trials where progression, not response, is the primary endpoint particularly if not all patients have measurable disease; whether or how to utilize newer imaging technologies such as [¹⁸F]-2-fluoro-2-deoxy-D-glucose (¹⁸F-FDG)-PET, PET-CT and MRI; how to handle assessment of lymph nodes; whether response confirmation is truly needed; and the applicability of RECIST in trials of targeted drugs. The RECIST 1.1 touches on all these points⁷.

Major changes in RECIST 1.1 include:

1. Number of lesions to be assessed: Based on evidence from numerous trial databases merged in to a data warehouse for analysis purposes⁶⁷, the number of target lesions is reduced from a maximum of 10 in RECIST to a maximum of 5 total and from 5 to 2 per one organ.
2. Assessment of pathological lymph nodes: Nodes with a short axis of ≥ 15 mm are considered measurable and assessable as target lesions⁶⁸. The short axis measurement should be included in the sum of lesions in calculation of tumor response. Nodes that shrink to < 10 mm short axis are considered normal.
3. Lytic bone lesions or mixed lytic-blastic lesions with identifiable soft tissue components that can be evaluated if the soft tissue component meets the definition of measurability. Totally blastic bone lesions are non-measurable.
4. Target lesions which are reported as “too small to measure” at follow-up should be assigned a default measurement of 5 mm if they are visible.
5. If isotropic reconstructions are performed on CT, measurements can be made on not only in trans-axial plane but also in a coronal or sagittal plane. Using the same plane of evaluation, the maximal diameter of each target lesion should always be measured at subsequent follow-up time points. Software tools that calculate the maximal diameter for a perimeter of tumor may be employed and may even reduce variability.

6. Confirmation of response is required for trials with response primary endpoint but is no longer required in randomized studies.
7. Disease progression is clarified in several aspects: in addition to the previous definition of progression in target disease of 20% increase in sum, a 5 mm absolute increase is required in order to guard against over calling PD when the total sum is very small. Furthermore, there is a stipulation what constitutes “unequivocal progression” of non-measurable/non-target disease.
8. The interpretation of ^{18}F -FDG-PET scan assessment is included in a section on detection of new lesions.
9. CT scans of PET-CT: If the CT performed as part of a PET-CT is of identical diagnostic quality to a diagnostic CT with i.v. and oral contrast, then the CT portion of the PET-CT can be used for RECIST measurements.
10. A lesion identified on a follow-up study in an anatomical location that was not scanned at baseline is considered a new lesion.

1.4 TUMOR MEASUREMENTS BEYOND WHO-CRITERIA AND RECIST

1.4.1 Waterfall analyses

The use of tumor measurements extends beyond RECIST. Increasingly, investigators use measurements in waterfall analyses and evaluate responses, progressions and simply change in tumor size as a continuous variable rather than categorizing the changes in tumor size^{40, 67} (Figure 4). A patient whose tumor shrinks by 25% has a different outcome than one whose tumor increases by 10%, yet both are considered by RECIST as having SD. There is not much difference between 32% shrinkage and 27% shrinkage, yet the former patient is classified as a PR and the latter is not PR but SD. Moreover, the ORR as an endpoint especially in phase II study is questioned and considered as a problem for evaluating molecular target drugs such as bevacizumab, cetuximab and so on. These drugs may be active even if they do not lead to tumor regression²⁰.

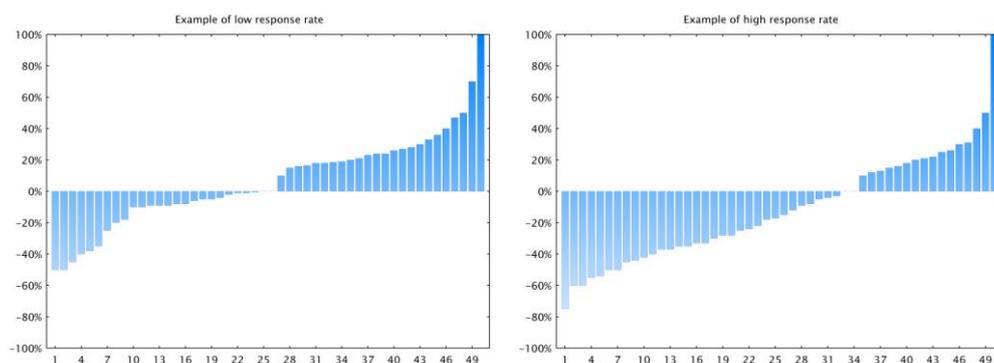


Figure 4. Example of Waterfall plots with low ORR (on the left) and with high ORR (on the right). The waterfall plots show the percent relative change in tumor size (y-axis) for each patient (x-axis). According to the distribution of plots, treatment effect can be estimated.

In this analysis, there are further advantages to treating size as a continuous variable⁶⁹. It is easy to incorporate covariates into the analysis, to formulate a simulation model for better planning for a phase III trial⁶⁹.

1.4.2 Functional Imaging Based Evaluation

There is no doubt that functional evaluation such as PET will provide a surrogate endpoint beyond tumor size for assessing the clinical efficacy of cancer drugs especially using new targeted agents. Though PET is costly, their availability is limited, and the sensitivity to detect changes in lesions less than 1 cm is uncertain⁷⁰, several studies have shown that ¹⁸F-FDG-PET is of value for evaluating targeted drugs not only

for imatinib (Gleevec) treatment of GIST metastases but other drugs as well ^{26, 27, 71}. Compared to one- and (RECIST) two-dimensional (WHO) measurements, SUV measurement has the highest reproducibility ¹⁵. Furthermore, the efficacy of ¹⁸F-FDG-PET has also shown promise for monitoring treatment response for earlier detection of responders (Figure 5). This may increase the possibility of shortening trial times ^{72, 73}. Early identification of responders and non-responders is important in order to select patients who might not experience any therapeutic benefit, to consider alternative therapies, and to avoid toxicity; early identification may also help to minimize the expense of clinical trials.

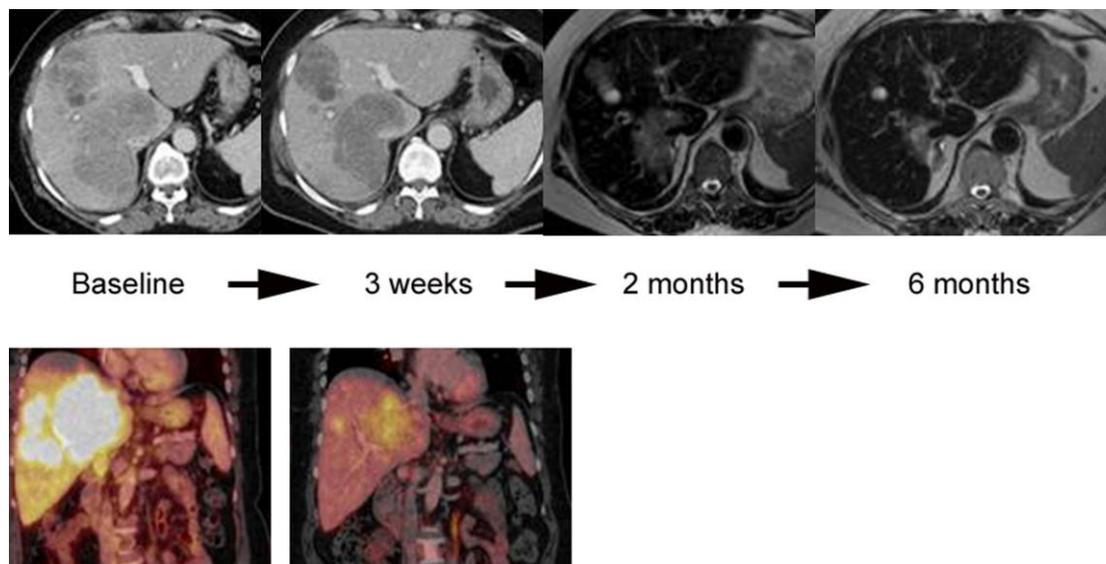


Figure 5. CT/MRI may delay detection of response compared to PET

A patient with multiple liver metastases from breast cancer treated with sunitinib (Sutent[®]). ¹⁸F-FDG-PET-CT obtained 3 weeks after therapy showed decreased ¹⁸F-FDG-uptake indicating response. On the other hand, it took 6 month to be categorized as PR on CT and MRI. (The patient had an adverse reaction to contrast agent for CT and therefore the follow-up modality was changed to MRI.)

Until widely accepted guidelines were introduced, PET could not gain acceptance as a tool for functional evaluation of therapy. EORTC and NCI PET guidelines and more recently PET Response Criteria in Solid Tumors (PERCIST) have been stipulated and recommended how ¹⁸F-FDG-PET should be performed and translated for response evaluation ⁷⁴⁻⁷⁶. In these guidelines, the SUV is recognized as valuable in the evaluation of ¹⁸F-FDG uptake in PET. However, the SUV is depending on several factors such as uptake time, the patient's body weight (which often changes during therapy), blood glucose levels, definition of Regions-Of-Interest (ROI), attenuation corrections, and the effects of metabolism by other tissues.

An advantage of PET/CT is that the CT-examination can be used for RECIST measurements if the CT exhibits the same diagnostic quality as a dedicated diagnostic CT (with i.v. and oral contrast medium) ⁷.

1.5 EMERGING IMAGE TECHNOLOGY AND CANCER DRUGS

In spite of remarkable progress in understanding molecular mechanisms in cancer cells, treatment response evaluation of solid tumors relies on changes in size assessed by imaging, mainly computed tomography (CT). Newer molecular targeted drugs have demonstrated an inherent limitation and unsuitability with “anatomical” tumor evaluation that only assesses the lesion size increase or decrease ^{20, 40, 77, 78}. The effect of these new target drugs demands the paradigm shift in evaluation,

Meanwhile, the widespread use of multi-detector computed tomography (MDCT) and other imaging innovations, such as PET-CT, MRI with functional capabilities such as dynamic contrast enhanced imaging and diffusion weighted image (DWI) ^{79, 80}, have led to concomitant need for modifications and up-date of RECIST (Figure 6).

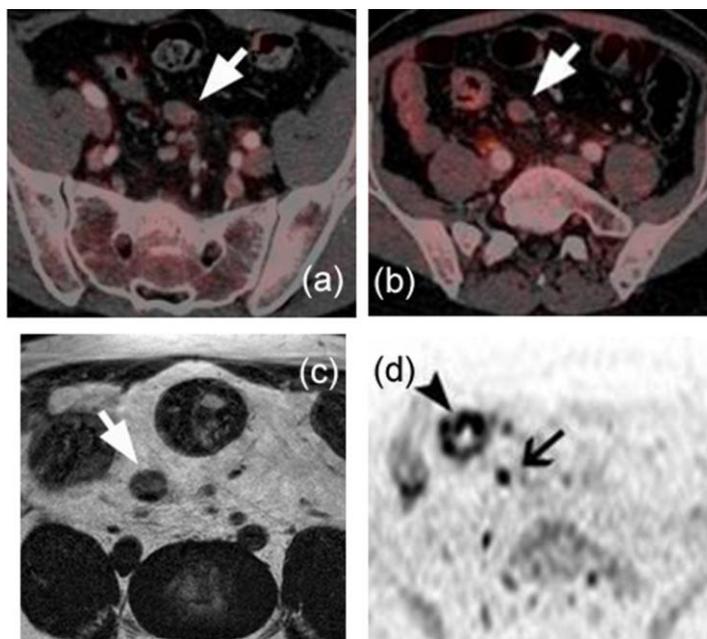


Figure 6. Lymph node metastasis in a patient with sigmoid colon cancer ¹⁸F-FDG-PET before (a) and after (b) treatment show only faint accumulation of ¹⁸F-FDG in a lymph node (white arrow). The T2-weighted image on MRI (c) shows heterogeneous higher intensity inside a node (arrow) that suggests pathological, possibly metastasis. DWI (b=500) highlights the node (black arrow) but it is unspecific without telling metastatic status. The primary lesion (arrowhead) in the sigmoid is also highlighted.

2 AIMS

To explore possible limitations with currently used image based tumor response criteria: WHO-criteria and RECIST, and to test if the tumor size change at the first response evaluation after initiation of chemotherapy correlates with outcome in patients with metastatic colorectal cancer and metastatic breast cancer.

Specific questions addressed for this purpose were the following:

2.1 STUDY I

RECIST does not specify the minimum number of target lesions. The aims of this study were whether the number of target lesions could be reduced, and whether the results still could be equally representative. If the number of target lesions could be reduced, how many target lesions would be needed to keep consistency.

2.2 STUDY II

What is the extent of inter- and intra-observer variations in RECIST and WHO-criteria based tumor response evaluation? What are the reasons for such variations?

2.3 STUDY III

Does the change in size of target lesions at the first follow-up CT after initiation of chemotherapy (1st change) correlate with PFS and OS in mCRC?

2.4 STUDY IV

Does the change in size of target lesions at the first follow-up CT after initiation of chemotherapy (1st change) correlate with OS in MBC?

3 MATERIALS AND METHODS

3.1 STUDY SUBJECTS

All studies were carried out retrospectively. Tumor size measurements and response definition were retrieved in the case recorded forms (CRFs) for study I, III and IV. In study II, tumor size measurement and BOR were reanalyzed independently by two board certified radiologists.

All studies were approved by the regional ethical committee.

3.2 DEFINITION OF OR ACCORDING TO RECIST AND WHO-CRITERIA

In study I, II, III and IV, response evaluation was conducted according to RECIST 1.0⁵.

CR means disappearance of all signs of tumor; PR means at least a 30% decrease in the sum of the largest diameters of the target lesions; PD means at least a 20% increase in the sum of the largest diameters of the target lesions or the appearance of new lesions or unequivocal progression of non-target lesions; SD means neither PR nor PD.

In addition to the RECIST evaluation, the perpendicular longest diameter of target lesion was obtained to allow a bi-dimensional evaluation according to WHO-criteria was conducted in study II. Different from RECIST, a patient was classified as PD if the sum of the products of the largest diameter (LD) and perpendicular LD increased 25% and more in one or more target lesion.

Objective response rate (ORR) was defined as the percentage of the sum of CR and PR (thus those who were considered responders) in relation to all patients: $(CR+PR)/\text{all patients}$.

3.3 STUDY I

A total of 99 patients enrolled in 16 phase II clinical trial studies at Karolinska University Hospital Solna with at least two CT-examinations amenable to measurements according to RECIST and who had completed their studies were eligible.

Target lesions were referred as 1L, 2L ... up to xL from the largest to the smallest. For 1L evaluation, only the largest; 1L, was chosen to assess the response. In the same fashion, for 2L evaluation, only the largest and the second largest; 1L+2L were chosen to assess the response. Thereafter, the discrepancy between full RECIST evaluation and

evaluation of fewer target lesions, in this study from seven down to one lesion, was analyzed.

3.4 STUDY II

Two board certified radiologists with experienced in oncological imaging retrospectively re-evaluated 39 patients' CT-examinations in nine phase II or III clinical trials for treatment of metastatic breast- and colorectal cancer according to RECIST and WHO-criteria. The radiologists independently selected and measured target lesions, assessed presence of new lesions or progression of non-target lesions and categorized treatment response. One radiologist repeated the procedure on two additional different occasions to examine intraobserver variation (Figure 7).

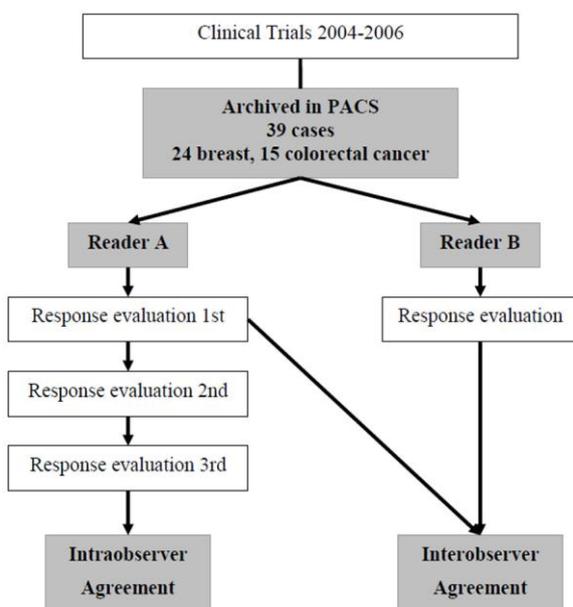


Figure 7. Scheme of study II (Reprinted from Acta Oncologica 2010⁸¹, copyright 2012, by permission of Informa Healthcare)

Kappa statistics was applied to examine inter- and intraobserver agreement. Kappa values above 0.81, 0.61 – 0.8, 0.41 - 0.60, 0.21 – 0.40, and 0-0.20 indicated almost perfect, substantial, moderate, fair and slight agreements, respectively⁸². Wilcoxon matched pairs test was used to compare the number of selected target lesions by readers. Friedman ANOVA by ranks was used to compare the number of selected target lesions among three repeated evaluations. A value of $p < 0.05$ was considered significant. Statistical analysis was carried out using StatXact 4 (CYTEL Software Corporation, Cambridge, MA, USA).

3.5 STUDY III AND IV

3.5.1 Definition of first change

The first change in tumor size (1st change) was calculated as the ratio of the sum of the LD of target lesions at the baseline study (baseline sum) and the sum of the LD of the same target lesions at the first follow-up CT (1st sum).

$$1^{\text{st}} \text{ change} = [(1^{\text{st}} \text{ sum}) - (\text{baseline sum})] / (\text{baseline sum})$$

Likewise, the 2nd change was calculated by the ratio of the difference of the sum of the LDs at the first and second follow-up studies in the study III.

3.5.2 Survival analyses

PFS and OS are the time from randomization to the date of any progression or death. In study III, the primary endpoints were the correlations between 1st change and OS and PFS. In study IV, the primary endpoints were the correlation between 1st change and OS.

Patients alive or progression-free at the last follow-up were censored.

Chi-square test or Fisher's exact test were applied to compare two independent groups. The Cox regression analysis was used to define prognostic variables and to estimate Hazard ratios (HRs) and 95% confidence limits (CL). The durations of PFS and OS were estimated by the Kaplan-Meier method. The log-rank test was used for comparison between groups.

3.5.2.1 *Guarantee-time bias and the landmark method*

The comparison of survival by response may be biased as patients need to live long enough to experience a response to treatment. Responders may survive longer than non-responders, not because of an effect of response on survival, but because response identifies patients with pretreatment characteristics that favor longer survival. Hence, the classification of patients as responders or non-responders is subject to "guarantee-time" bias, so that patients who die early before they have had a change of responding to treatment have a major negative impact upon the survival estimates of the group of patients classified as non-responders. The presence of guarantee-time bias can be eliminated by the "landmark method". This essentially consists of setting the time origin for survival after a landmark period long enough for all responses to have been observed^{1, 83, 84}.

For example, Figure 8 shows the survival for responders versus non-responders without landmark method. Responders have a median survival which is more than twice as long

as the median survival for the non-responders. The difference between the two curves is, however, due to the four non-responders who died early, before showing response or non-response to the treatment. The bias caused by the guarantee time is clearly visible in this example (Figure 8) ¹.

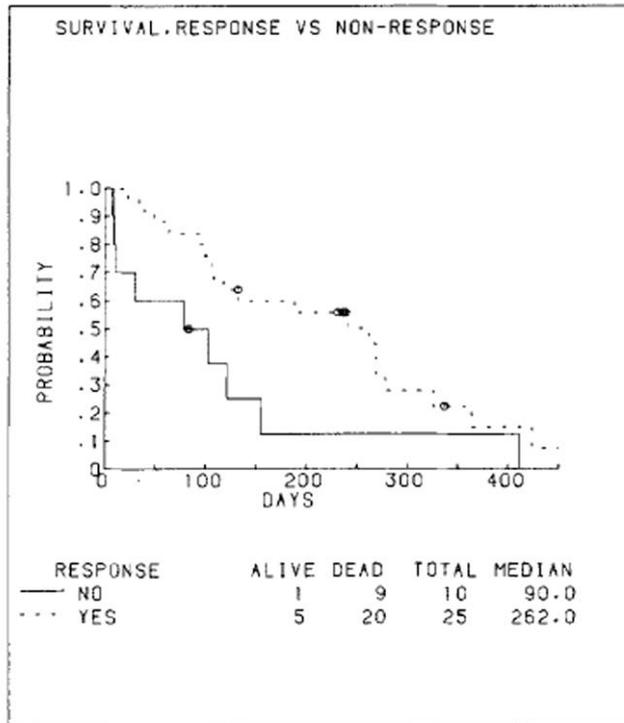


Figure 8. Evaluation of survival for responders vs. non-responders without landmark method.

(reprinted from Journal of Clinical Oncology 1983, Anderson et al ¹, by permission of American Society of Clinical Oncology)

The landmark time analysis in which survival from the landmark time is evaluated by response status is shown in Figure 9.

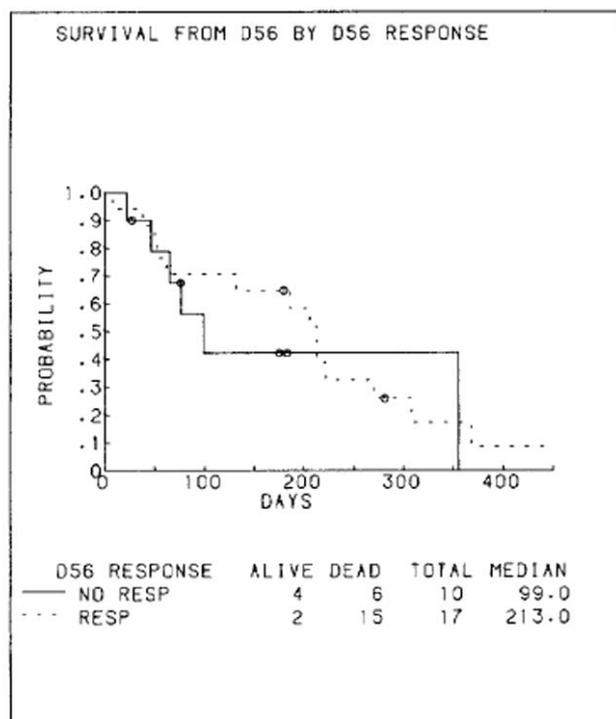


Figure 9. Evaluation of survival for responders vs. non-responders with landmark method.

(reprinted from Journal of Clinical Oncology 1983, Anderson et al ¹, by permission of American Society of Clinical Oncology)

In this study, patients who died or relapsed prior to the landmark period have been excluded from the analysis. No effect of response is apparent when the guarantee-time (landmark time) bias has been removed (Figure 9) ¹.

Patients who survive the landmark period are classified as responders or non-responders, and their survival time after the end of the landmark period can be compared without bias. Using this method, either patients who die before the landmark time point are excluded from the analysis, or the landmark time is subtracted from PFS or OS if patients who live longer than the landmark time point. The landmark method was used to compensate for longer guarantee-time for responders and was applied in the PFS and OS analyses ⁸³.

3.5.2.2 *The intra-class correlation coefficient (ICC)*

The intra-class correlation coefficient (ICC) measures the strength of agreement between repeated measurements by assessing the proportion of the between-patient variance (the square of the standard deviation) versus the total variance, comprising the sum of both the within and between variation. The ICC ranges from 0.00 to 1.00; with 1.00 representing perfect repeatability, hence high reliability. The ICC is the most commonly used method for assessing reliability with continuous data as well as categorical data. A value of at least 0.90 is often recommended if the measurements of concern are to be used for evaluating future patients for which therapeutic decisions are to be made⁸⁵.

To assess the reliability of 1st change values, an independent board certified radiologist retrospectively selected tumors and measured the baseline sum and the 1st sum in study III and IV. The radiologist had no information about the original radiological response assessments or patient outcomes.

3.5.3 Statistical computations

A value of $p < 0.05$ was considered significant. Statistical computations were performed with StatXact 4 (CYTEL Software Corporation, Cambridge, MA, USA) and STATISTICA ver. 9 (StatSoft Inc., Tulsa, OK, USA).

3.5.4 Patients in Study III

The patients were enrolled in the multi-center randomized phase III Nordic VI trial ($N=567$) comparing irinotecan with either the Nordic bolus 5-fluorouracil (5-FU) and folinic acid schedule (FLIRI) or the de Gramont schedule (Lv5FU2-IRI)⁸⁶. Both FLIRI and Lv5FU2-IRI treatments were repeated every 2 weeks until disease progression or unacceptable toxicity. CT scans of the chest and the abdomen were obtained at baseline within 3 weeks prior to randomization and repeated after every 4 cycles (every 8 weeks). Response evaluation was conducted according to RECIST 1.0.

According to the original study, there was no significant difference in PFS, median 9 months ($p=0.4$), or OS, median 19 months ($p=0.4$) in the two groups; however, ORR were more common using LvFU2-IRI compared to FLIRI (49% vs. 35%, $p=0.001$)⁸⁶.

A total of 506 patients were eligible for the study. Patient characteristics are shown Table 2.

Table 2. Patient Characteristics in Nordic VI (n =506)

Characteristic	
Age (years)	
Mean	61
Median	62
Range	29 - 76
Sex	
Male	314 (62%)
Organ involved	
liver	405 (40%)
lung	162 (16%)
lymph nodes	119 (12%)
others	336 (33%)
No. of organs involved	
1	244 (48%)
2	162 (32%)
3+	100 (20%)
Overall Response / No. operated	
CR	23 (5%)/5 (22%)
PR	208 (41%)/11 (5%)
SD	217 (43%)/10 (5%)
SD4 ^a	159/8
SD2 ^a	50/2
PD	58 (11%)/0 (0%)
Reason for PD at the 1 st follow-up (n=58)	
Size only	24 (41%)
New, non-target	33 (57%)
Other clinical reasons	1(2%)
Organ for new metastases, non-target progression	
liver	23
lung	5
bone	3
others	6

^a the numbers of SD4 and SD2 were given for those who underwent the 2nd follow-up after 4 months (n=432).

New, non-target: appearance of new lesion or progression of non-target lesion

3.5.5 Patients in Study IV

Totally, 287 patients with metastatic breast cancer were enrolled in a multi-center randomized phase III trial comparing a combination of epirubicin (Farmorubicin[®]) and paclitaxel (Taxol[®]) alone (ET) or in combination with capecitabine (Xeloda[®], TEX)⁸⁷. Study subjects with morphologically confirmed loco-regional inoperable or disseminated carcinoma were randomized unless they had received treatment with an anthracycline, a taxane or 5-FU within one year before study entry. Previous endocrine treatment for advanced disease in patients with hormone receptor positive breast cancer was allowed. Patients with known brain metastases or other malignancies within the last five years were excluded.

In this trial, chemotherapy was repeated every 3rd week until disease progression or severe adverse effect. Baseline tumor measurement was obtained by CT, MRI or palpation 4 weeks prior to randomization. Tumor measurement was repeated after every 3 cycles (every 9 weeks). Response evaluation was conducted according to RECIST.

A total of 233 patients were eligible for the study. Cases with missing measurements in one of the two evaluations (n=26) or only non-measurable lesions (n=28) were excluded. Patient characteristics are shown in Table 3.

Table 3. Patient Characteristics in TEX (n = 233)

Characteristic		
Age at entry	Mean	54
	Range	29 - 74
No. of lesions/ patient	Mean	2.8
	Median	3
	Range	1 - 9
No. of target lesions evaluated	Liver	262 (39.9%)
	Lymph node	235 (35.8%)
	Lung	86 (13.1%)
	Others	74 (11.3%)
	Total	657
Treatment arm	TEX	117
	ET	116
Previous adjuvant treatment	Yes	116
	No	117

4 RESULTS

4.1 STUDY I

The number of discordant results in response categories increased gradually from 0% for 4 lesions to 8% for one lesion assessments when all cases were considered. When only patients with ≥ 5 measurable lesions (53 cases) were assessed, the discordant rate ranged from 0% for 4L to 15.1 % (Table 4).

Table 4. Number of cases with discordant results between the measurement of fewer lesions and measurement of all target lesions

Type of assessment	1L	2L	3L	4L, 5L, 6L, 7L
Number of "mistakes" (discordant cases) when all 99 patients are considered	8	6	4	0
Percentage of discordant cases in all 99 cases	8%	6%	4%	0%
Percentage of discordant cases in patients with ≥ 5 lesions (53 cases)	15.1%	11.3%	7.5%	0%

(Reprinted from The British Journal of Radiology 2009⁸⁸, copyright 2012, by permission of British Institute of Radiology)

The mean number of measurable lesions among those with discrepancies was 7.1. This was significantly higher than those demonstrating concordance when measuring fewer target lesions and all measurable lesions (4.1 lesions; $p < 0.05$).

Figure 10 shows the percentage of discordances based on the available number of target lesions and the number of target lesions chosen.

Measuring less than four target lesions increased discrepancies in terms of BOR when more than five target lesions are present.

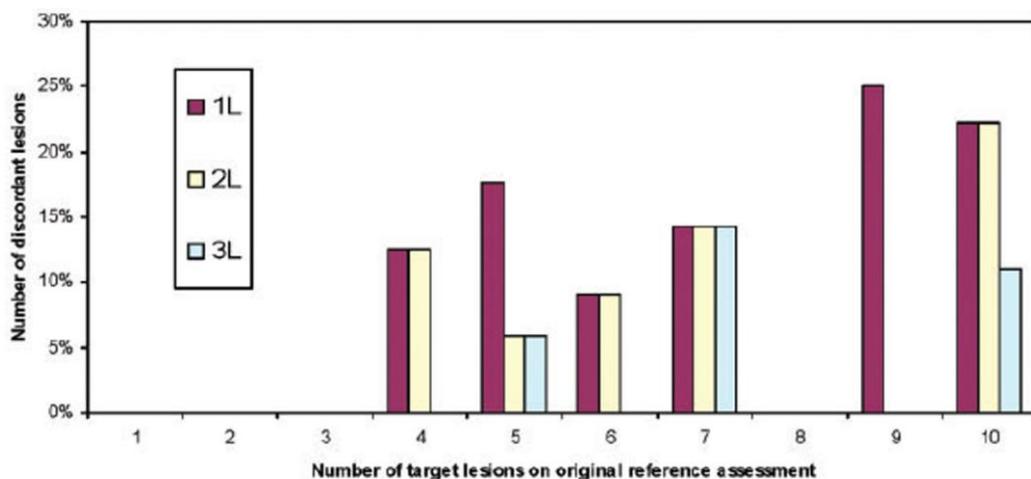


Figure 10. Number of discordant cases based on available target lesions. (Reprinted from The British Journal of Radiology 2009⁸⁸, copyright 2012, by permission of British Institute of Radiology)

4.2 STUDY II

The number of CT-examinations followed for response evaluation was different between the readers because of judging the same patient as PD at different time points (Figure 11). The numbers of selected target lesions were also different between the readers (Figure 12).

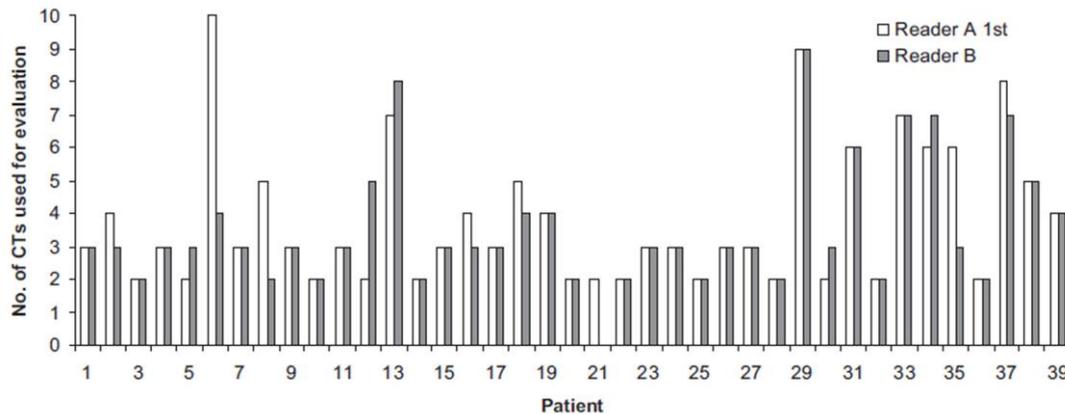


Figure 11. Number of CT-examinations used for evaluations by reader A 1st and reader B. (Reprinted from Acta Oncologica 2010⁸¹, copyright 2012, by permission of Informa Healthcare)

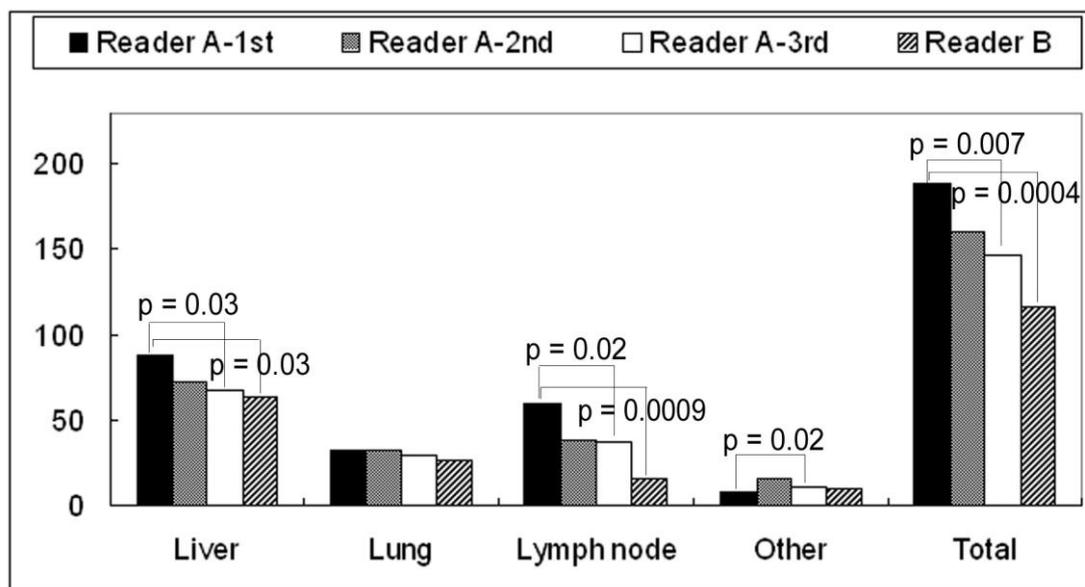


Figure 12. Number of target lesions selected by reader A and B. (Reprinted from Acta Oncologica 2010⁸¹, copyright 2012, by permission of Informa Healthcare)

As shown in Figure 13, interobserver agreement was moderate and generally lower than the intraobserver agreement, which tended to be substantial to perfect. RR according to RECIST was 33% by reader A and 21% by reader B. RR according to bi-dimensional measurements simulating WHO-criteria was 33% by reader B and 23% by reader B.

Interobserver agreement in the detection of new lesion(s) and progression of non-target lesions was moderate, $\kappa = 0.50$ (0.28–0.73). Even among five patients whose target lesions were identical between the two readers, patients were judged differently mainly because inconsistency in interpretation of new lesions and progression of non-target lesions. In seven out of 39 patients, there was a clear discordance regarding evaluation of PD. Intraobserver agreements in the detection of new lesion(s) or progression of non-target lesions ranged from substantial to perfect.

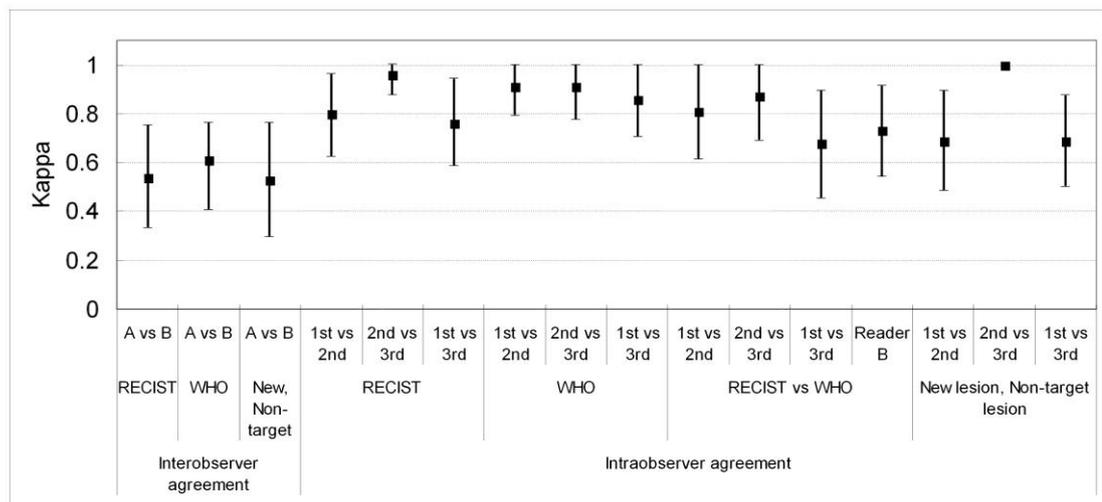


Figure 13. Non-weighted kappa coefficient value and corresponding 95% confidence interval for agreement. (Reprinted from Acta Oncologica 2010⁸¹, copyright 2012, by permission of Informa Healthcare)

It is concluded that radiological tumor response evaluation according to RECIST and bi-dimensional tumor measurement according to WHO-criteria is subjected to a considerable inter- and intraobserver variability.

4.3 STUDY III

The HRs for PFS and OS decreased with 1st change. A decrease between 10% to <30%, albeit RECIST does not regard this as a partial response, was a positive prognostic factor for PFS and OS. Patients who had new lesions or unequivocal progression of non-measurable lesions had a worse prognosis than those with only an increase in size of >20%.

4.3.1 Relationship between 1st change and OS in mCRC

The 1st change used as a continuous variable correlated strongly with OS. A decrease in size by 50%, or more, was the most positive prognostic factor for OS. In the absence of a new lesion or unequivocal progression of non-target lesions, even an increase of $\geq 20\%$, considered as PD by RECIST, was not statistically significantly associated with impaired OS vs. an increase of <10%. Appearance of a new lesion or progression of non-target lesion(s), irrespective of the change in diameters of the target lesions, was the most negative prognostic factor for OS (Table 5).

Table 5. Cox proportional hazards multiple regression model on overall survival in 506 patients with 327 deaths

Covariate 1st change	No. of patients	<i>p</i>	HR	95% CL	OR
Continuous*	506 (327)	<0.001	2.01	1.75 – 2.31	-
Categorized					
New, non-target	33 (29)	<0.001	3.77	2.08 – 6.83	PD
Increase ≥20%	26 (22)	0.611	1.18	0.63 – 2.20	
≥10 - <20%	15 (10)	0.722	1.15	0.53 – 2.49	SD
>0 - <10%	25 (18)	-	1	-	
0 – decrease <10%	104 (69)	0.310	0.76	0.45 – 1.28	
Decrease ≥10 - <20%	95 (57)	0.031	0.56	0.33 – 0.95	
≥20 - <30%	57 (31)	0.002	0.39	0.22 – 0.70	
≥30 - <40%	68 (47)	0.046	0.57	0.33 – 0.99	PR
≥40 - <50%	40 (25)	0.010	0.45	0.24 – 0.82	
≥50%	43 (19)	<0.001	0.25	0.13 – 0.49	

* 1st change as a continuous valuable. Patients with new lesions or progression in non-target lesion at the 1st follow-up study converted into an increase of 1.0.

Number in parenthesis indicates the number of deaths.

HR: hazard ratio

95% CL: 95% confidence limits

OR: overall response if based upon the 1st change only

New, non-target: appearance of new lesion or progression of non-target lesion

Based on the results of the Cox regression model, patients were regrouped into four groups: (i) having new lesion/unequivocal non-target lesion progression, (ii) any increase to <10% decrease, (iii) $\geq 10\%$ to <50% decrease or (iv) a decrease of $\geq 50\%$. The difference in OS was statistically significant among these four groups (landmark method, log-rank test, $p < 0.00001$) (Figure 14). The four RECIST categories could similarly discriminate OS (Figure 15).

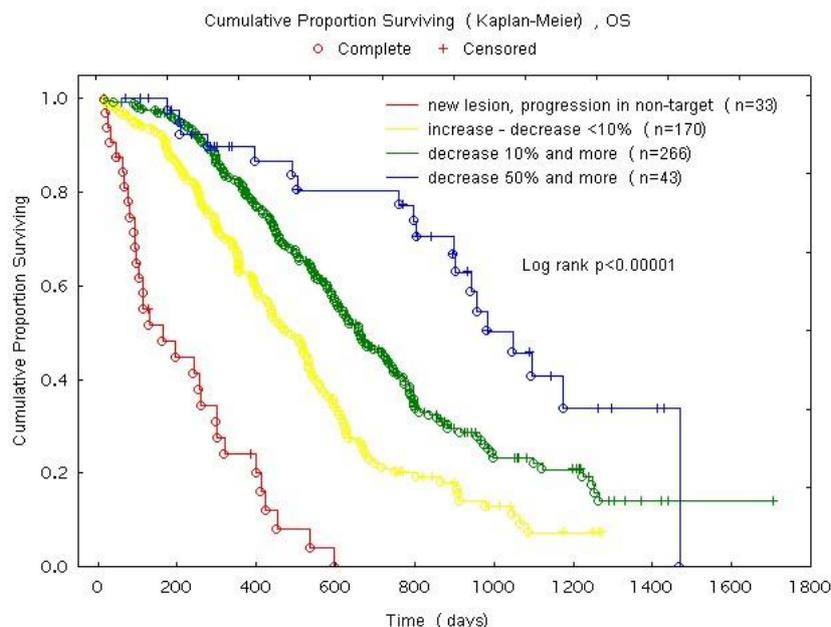


Figure 14. Kaplan-Meier curves of overall survival according to first change values

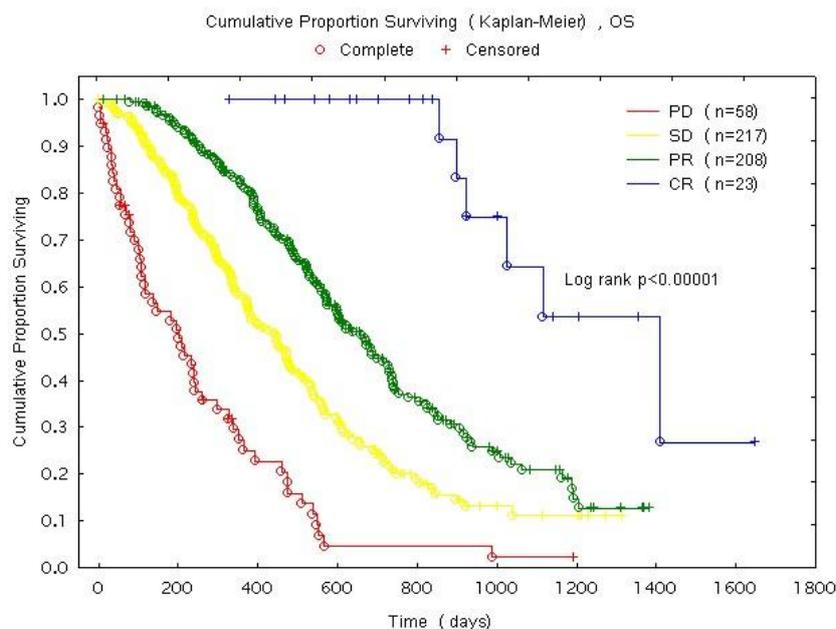


Figure 15. Kaplan-Meier curves of overall survival according to RECIST

4.3.2 Relationship between 1st change and PFS in mCRC

The PFS increased as the 1st change value decreased. A patient whose tumor shrank more survived longer without progression. A decrease in size by $\geq 50\%$ was again the most significant positive prognostic factor of PFS (Table 6).

Table 6. Cox proportional hazards multiple regression model on progression-free survival in 506 patients with 456 events

Covariate 1 st change	No. of patients	<i>p</i>	HR	95% CL	OR
Continuous*	506 (456)	<0.001	2.21	1.97 – 2.49	
Categorized					
New, non-target	33 (33)	<0.001	12.21	6.75 – 22.09	PD
Increase $\geq 20\%$	26 (25)	<0.001	3.51	1.98 – 6.23	
$\geq 10\% - <20\%$	15 (14)	0.413	0.76	0.39 – 1.47	
$>0 - <10\%$	25 (24)	-	1	-	
0 ~ decrease $<10\%$	104 (92)	0.167	0.73	0.46 – 1.14	SD
Decrease $\geq 10 - <20\%$	95 (82)	0.013	0.56	0.36 – 0.89	
$\geq 20 - <30\%$	57 (51)	0.004	0.49	0.30 – 0.80	
$\geq 30 - <40\%$	68 (66)	0.038	0.61	0.38 – 0.97	
$\geq 40 - <50\%$	40 (34)	<0.001	0.38	0.23 – 0.64	PR
$\geq 50\%$	43 (35)	<0.001	0.35	0.21 – 0.59	

* 1st change as a continuous valuable. Patients with new lesions or progression in non-target lesion at the 1st follow-up study converted into an increase of 1.0.

Number in parenthesis indicates the number of events.

HR: hazard ratio

95% CL: 95% confidence limits

OR: overall response if based upon 1st change only

New, non-target: appearance of new lesion or progression of non-target lesion

Patients were regrouped according to the 1st change values: (i) having new lesion/non-target lesion progression, (ii) any increase to a <10% decrease, (iii) $\geq 10\%$ to <50% decrease or (iv) $\geq 50\%$ decrease. The difference in PFS among these four groups was statistically significant (landmark method, log-rank test, $p < 0.00001$) (Figure 16). The same was seen for the RECIST categories (Figure 17).

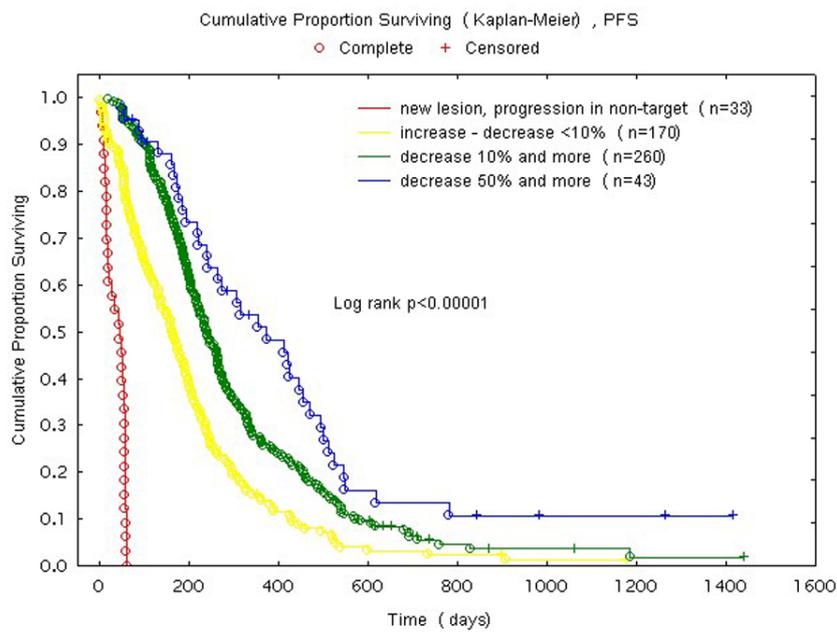


Figure 16. Kaplan-Meier curves of progression-free survival according to first change values

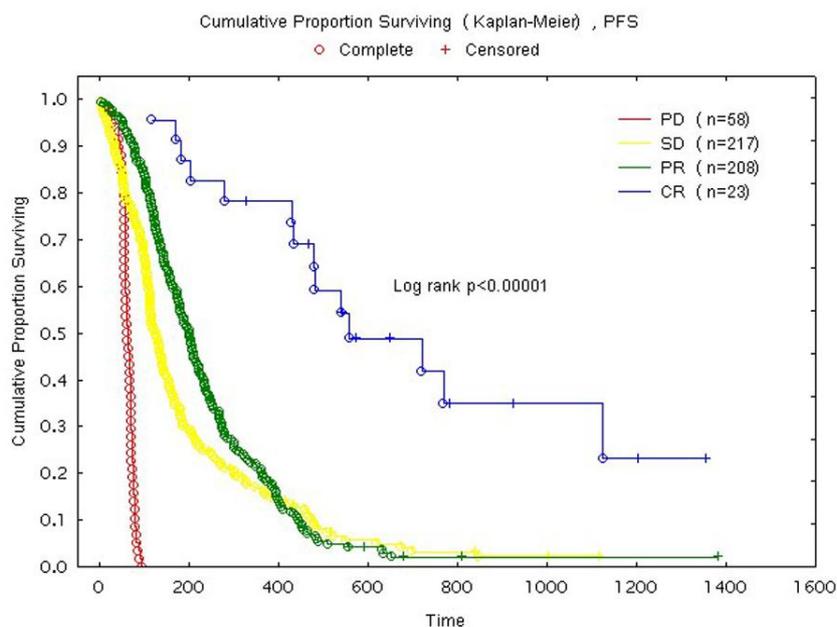


Figure 17. Kaplan-Meier curves of progression-free survival according to RECIST

4.3.3 The 2nd follow-up examination

At the 1st evaluation, 283 patients were evaluated as SD. Of these, 43 patients (15%) had disease progression before 4 months (designated SD2).

Patients with SD4 had significantly longer OS compared to those with SD2 (median 497 vs. 358 days, log-rank test, $p=0.00742$). The mean 1st change values for SD2 and SD4 were -0.12 (SD 0.19) and -0.09 (SD 0.13), respectively, without any statistically significant difference ($p=0.13$). However, the chance of being SD4 was significantly higher if a patient's 1st change showed a decrease of $\geq 10\%$ ($p=0.00658$) (Figure 18).

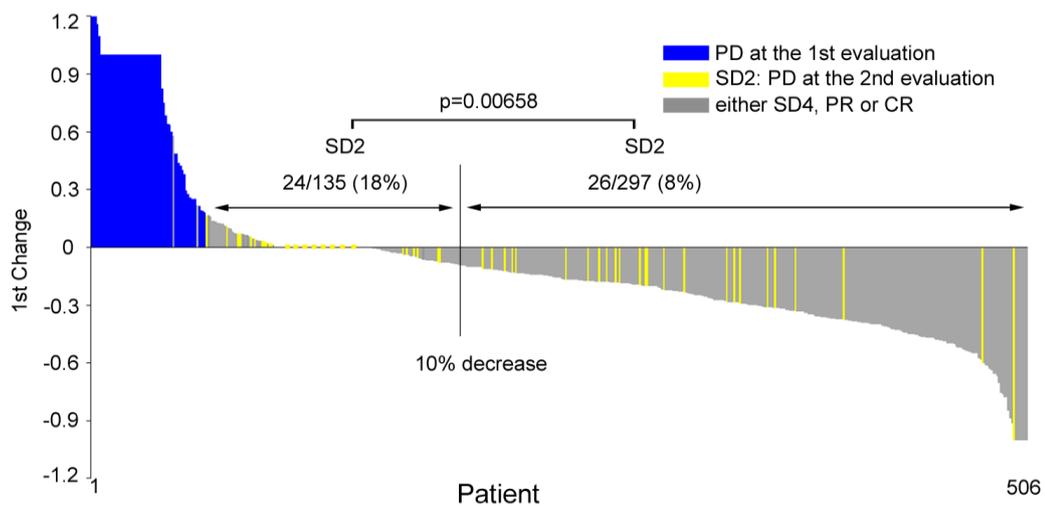


Figure 18. Waterfall plot of the first change. Patient with new lesion at the first follow-up was converted into 1.0 (an increase of 100 %).

Similar to the strong predictive capacity of PFS and OS by the 1st change from baseline to 2 months, the appearance of a new lesion or progression of non-target lesion(s) at the 2nd change was a single significant negative variable and a decrease by $\geq 50\%$ was a single significant positive variable for both PFS and OS (Table 7).

Table 7. Prognostic variables by Cox proportional hazards multiple regression model on PFS and OS at the second time (n=432).

Covariate (2nd change)	No. of patients*	<i>p</i>	HR	95% CL
PFS				
New, non-target	28 (28)	< 0.001	17.85	9.78 – 32.58
Decrease $\geq 50\%$	41 (35)	0.008	0.50	0.30 – 0.84
OS				
New, non-target	28 (26)	0.014	2.01	1.15 – 3.52
Decrease $\geq 50\%$	41 (17)	<0.001	0.25	0.13 – 0.47

The group which showed no change to up to a 10% increase at the second follow-up was referred to as a reference group (HR=1.0).

Other variables; from any increase to decrease <50%, were not significantly different from 1.0.

* Number in parenthesis indicates the number of events.

HR: hazard ratio

95% CL: 95% confidence limits

New, non-target: appearance of new lesion or progression of non-target lesion

4.4 STUDY IV

An increase of the sum of metastatic lesions by more than 10% or the appearance of new lesions or progression of non-target lesions (increase >10%/new/non-target) predicted a shorter OS by univariable regression analysis (HR 4.29, 95% CI 2.44 – 7.53, $p < 0.001$). After adjustment for previous adjuvant treatment and the treatment given within the frame of the randomized trial, OS was still significantly shorter in the group of patients with an increase of the sum of metastatic lesions by more than 10% or the appearance of new lesions or progression of non-target lesions (HR 4.58, 95% CI 2.86 – 7.33, $p < 0.001$).

4.4.1 The extent of 1st change in MBC

The median time from randomization to the first evaluation was 8.3 weeks (SD 1.6 weeks). The extent of the first change is shown in Figure 19. Twenty-three (10%) patients had a new lesion or progression of non-target lesions at this point; eight of these in combination with more than a 20% increase of target lesions; i.e. cut-off point for PD according to RECIST. No patient had solely more than a 20% increase of target lesions. Survival longer than the median OS (blue column in Figure 19) was seen both in patients who had at least a 30% decrease (PR for RECIST), and among those with a smaller decrease (10 – 30%), or no change (<10% increase or decrease) (SD for RECIST).

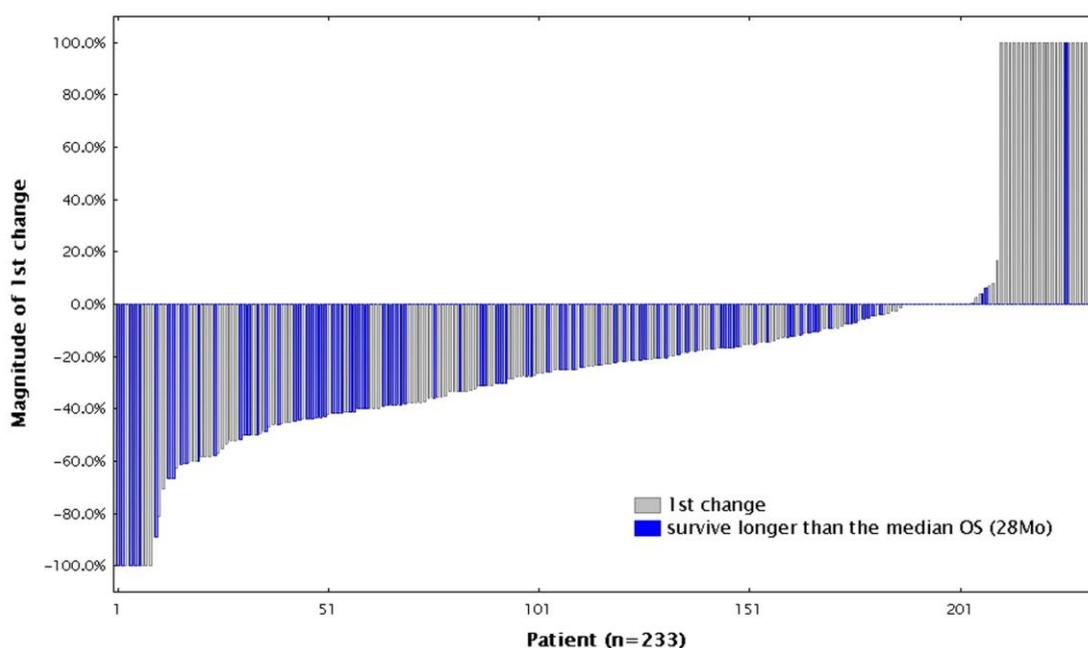


Figure 19. Waterfall plot of the first change of tumor size.

4.4.2 Relationship between 1st change and OS in MBC

Univariate regression analyses showed that the group of patients with more than a 10% increase in tumor size or appearance of new lesions or progression of non-target lesions at the first evaluation (increase >10%/new/non-target) had a significantly increased hazard of death (HR 4.29, 95% CI: 2.44 – 7.53, p<0.001). Groups with no change (\pm 10%) and decrease in size by more than 10% had a similar outcome (Table 8). According to RECIST, no response (SD or PD) increased the risk for an unfavorable outcome (HR 1.97, 95% CI: 1.43 – 2.71, p<0.001, Table 8).

Table 8. Univariable Cox regression analyses on overall survival (OS) in 233 patients with 158 events

	No. (%)	HR	95% CI	p-value
Change of size at the first response evaluation				
Increase >10%/new/non-target ^(a)	24 (10)	4.29	2.44 – 7.53	<0.001
No change \pm 10%	42 (18)	1	-	-
Decrease >10%	167 (72)	0.86	-0.56 – 0.25	0.45
Overall response				
Responders (CR, PR)	138 (59)	1	-	-
Non-responders (SD, PD)	95 (41)	1.97	1.43 – 2.71	< 0.001

^(a) new/non-target: appearance of new lesion or progression of non-target lesion

There was a significant difference in OS between patients with an increase >10% /new/non-target and those without such findings (log-rank $\chi^2=20.4$, $p<0.001$; Figure 20). Corresponding differences were found between responders and non-responders according to RECIST (log-rank $\chi^2=15.8$, $p<0.001$, Figure 21).

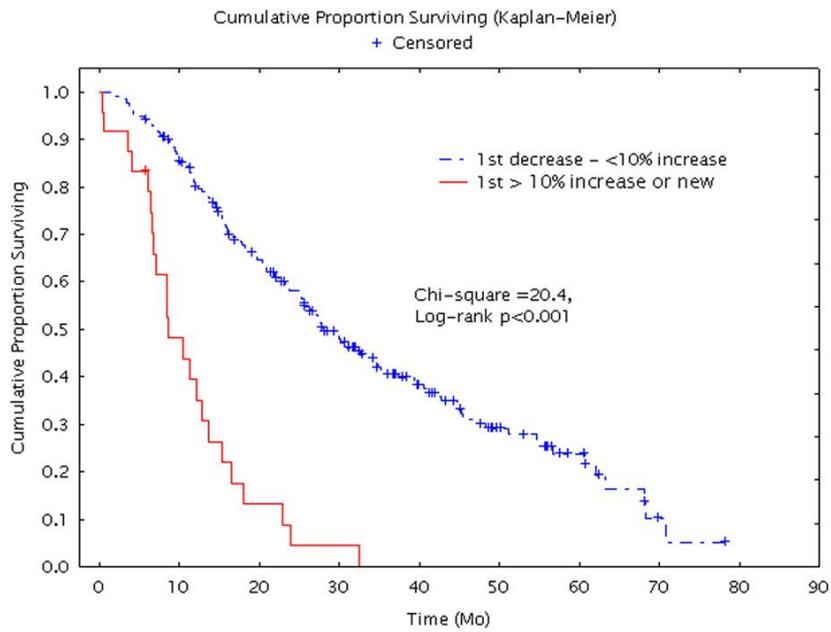


Figure 20. Kaplan-Meier curves of overall survival according to the first change

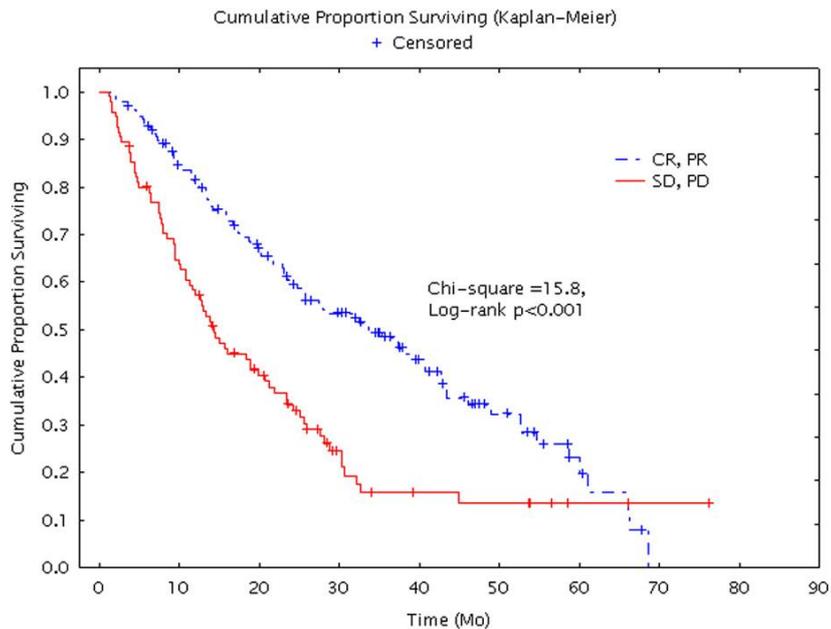


Figure 21. Kaplan-Meier curves of overall survival according to RECIST

4.4.3 The Cox regression analysis of OS in MBC

Clinical data from the TEX trial indicated that previous adjuvant treatment reduced the chance to respond to the study treatment⁸⁷. Adjusted for adjuvant treatment status and study treatment arm (ET vs. TEX) an increase >10%/new/non-target lesions was found to be an independent prognostic factor for OS (HR 4.58, 95% CI: 2.86 – 7.33, p<0.001, Table 9).

Table 9. Multivariate Cox regression model adjusting the first change of size to previous adjuvant treatment and study treatment arms in terms of overall survival (OS)

	HR	95% CI	<i>p</i>
Increase >10%/new/non-target ^(a)	4.58	2.86 – 7.33	< 0.001
Previous adjuvant treatment (None = 1)	1.40	1.02 – 1.92	0.04
Study treatment (ET=1.00)	0.92	0.67 – 1.26	0.59

^(a) new/non-target: appearance of new lesion or progression of non-target lesion

5 DISCUSSION

5.1 TO MEASURE OR NOT TO MEASURE, THAT IS THE PROBLEM

In order to allow adequate evaluation of many various anti-neoplastic drugs examined in clinical trials, quantitative information is required. The increasing demand to objectively verify the treatment response of new cancer agents has resulted in an increased use of medical imaging, especially CT. WHO-criteria and RECIST (1.0 and 1.1) were set up in order to convert the analogous information provided by the radiological studies into the digital information in a strict and standardized way^{3-5, 7}. The digital information, such as measurement values and changes before and after initiation of cancer agents appears to be very precise and objective, however the crucial, but subjective steps, when a lesion is selected and thereafter measured have surprisingly little been questioned.

In this thesis various sources of measurement variability in the process of tumor response evaluation has been revealed and described. These discordances can result in considerable discrepancies in response evaluation using both WHO-criteria and RECIST despite seemingly standardized conditions on how to carry out the evaluations. The possible negative effects of these findings on the reliability of clinical trials are obvious.

The significant variability between readers in the selection, in the measurement of the lesions as well as in tracking new lesions and in the attention of non-target lesions was shown in study II⁸¹. To minimize diversity in selection of target lesions, reducing the number of target-lesions might be suggested as a possible solution. However as shown in study I as well as several previous retrospective studies, reducing the number has negative impact on consistency^{17, 18, 88}. This is because the contribution of a single lesion to the response evaluation increases when the number of lesions is reduced. It is also doubtful that only one or two arbitrary selected lesions in a patient with multiple metastases can represent the patient's true response.

In our radiology department, one dedicated radiologist with experience in oncological imaging takes responsibility of response evaluation in each clinical trial. This one radiologist to one trial system ensures a more consistent evaluation, thus more reliable result⁸⁹. Independent central review (ICR) is a way to enable more objective, reproducible and independent evaluation of results compared to the response evaluation carried out in the daily workflow^{89, 90}.

Moreover, considerable advances in medical imaging technology such as automated CT volumetric measurement and computer aided detection (CAD) may help to improve accuracy in the measurement and to reduce intra- and inter-observer inconsistencies^{12-14, 91-93}. Alternative measurement values, for example SUV based on the PET, has potentially higher reproducibility than tumor size measurement on CT¹⁵.

Even though all of those inconsistencies in measurement were minimized, it is still questionable how much the tumor measurement itself correlates with a patient's prognosis. In studies II, III and IV, the main reason for PD was not size increase of target lesion but the appearance of new lesion and/or progression of non-target lesion. In addition, study III indicated a marked difference in prognosis between patients with or without a new lesion or unequivocal progression of non-target lesion(s) in patients with mCRC. To our knowledge, this has not been reported before, or incorporated in the simulation model⁴². Since there was a complete overlap between patients with more than 20% increase and the development of new lesions and progression of non-target lesions in study IV, we were not able to find differences in survival between PD with and without development of new lesions and progression of non-target lesions in patients with MBC. Yet, the presence of a new lesion or progression of non-target lesions was highlighted as potentially more critical for survival as any change in size in study IV. On the basis of our studies, one may speculate that the appearance of new lesions or progression of non-target lesions could mean that the tumor disease is biologically more aggressive than when the progression is only due to an increase in the size of target lesions.

5.2 CLINICAL IMPLICATIONS, WHAT MATTERS MOST?

Few physicians and researchers have questioned the relevance of thresholds or categorization for response according to RECIST. Instead, the value of functional imaging for early response prediction has been the focus, with contradictory results²⁵⁻²⁷. The relevance of early changes in size has not been well explored. The initial change in tumor size in mCRC (study III) and MBC (study IV) correlates with patient prognosis in terms of PFS (study III) and OS (study III and IV). This has previously never been shown besides for a modeling study to support end-of-phase II decision and design for phase III studies⁴². On the basis of studies III and IV, the comparison of cytotoxic treatments in a trial can be achieved more rapidly by the 1st change approach than waiting for BOR using RECIST.

Furthermore, these studies question the currently adopted cut-off values to evaluate tumor response. The cut-off values used by RECIST between PD and SD on the one hand and between SD and PR on the other do not seem to be optimally chosen. The chosen cut-off values should be reappraised from a clinical standpoint ⁶ and not only from the extent of measurement errors ^{2, 5, 14, 15, 63, 64, 94}.

The BOR and ORR have been used as a surrogate outcome that is likely to predict clinical benefit for long time ^{39, 95}. However, the value of the BOR and ORR as a surrogate indicator has been debated. Studies of different tumor types, particularly colorectal cancer (CRC) and MBC, have failed to show a strong correlation between a BOR and prolonged survival^{19, 21, 23, 83}. A survival benefit has been shown not only for patients who had a complete or partial response (CR+PR) but also for those who maintained SD for at least 4 months; the latter patients did not satisfy WHO-criteria requirements for PR ¹⁹. RECIST's thresholds were also recently questioned in a mCRC trial including the epidermal growth factor receptor (EGFR) inhibitor cetuximab ^{22, 96} and in a MBC trial including the EGFR inhibitor bevacizumab ²⁴.

Even other surrogate outcomes, PFS or TTP could not demonstrate as a good surrogate for OS in a first-line randomized clinical trials that compared an anthracycline with a taxane ²³ or in a phase III randomized trial that compared chemotherapy with or without bevacizumab ²⁴.

Why these surrogate outcomes do not correlate with OS? Besides thresholds setting, it should be questioned why dichotomization or categorization of response is necessary. Categorization often simplifies data analysis and interpretation. It is, however, obvious that once tumor size changes are dichotomized or categorized, e.g. responder and non-responder, it loses valuable information and statistical efficiency than continuous variable^{40, 69}. As shown study III, the first change as continuous variable might correlate better with OS.

It is crucial to find a surrogate outcome which correlates with OS. That may lead better planning for phase II and III trials, reduce the risk of failure of clinical trials, especially phase III trials and eventually lead reduction in drug developing costs ²⁰. We hope the first change described in this thesis can be an effective surrogate for OS and help to deliver effective treatments to patients.

5.3 LIMITATIONS

There are several limitations in the studies that constitute this thesis. First, all studies were conducted retrospectively. Second, RECIST 1.1 was not used in any study since it is relatively new, introduced in 2009, and there are only about 15 published articles that adopted RECIST 1.1 so far. We cannot tell that issues addressed in study I and II are still valid in RECIST 1.1. However, the impact of initial tumor size change and appearance of new lesion/progression of non-target lesion to patient's prognosis are of importance regardless which criteria is adopted.

Study I included a considerable number of patients while the mean number of measurable lesions per patient was limited. Larger studies with also more measurable lesions per patients are needed.

Study II is based on a restricted number of observations and on a heterogeneous patient group why it must be regarded as a pilot study. Ideally, it should be repeated in larger cohorts and at other institutions. Nevertheless and especially considering these limitations, the inconsistency of how lesions are selected, measured and interpreted between different readers remains an important finding.

Tumor measurement for WHO-criteria was actually "RECIST" based measurement, namely the perpendicular longest diameter was obtained for a lesion which was selected as a target lesion according to RECIST. This might cause underestimation of the inconsistency because there is no strict stipulation of number of target lesion or measurable lesion in WHO-criteria.

A limitation of studies III and IV is that only one time point defined by trial protocol for the follow-up study was analyzed (8 weeks for mCRC and MBC). We don't know if this time point is optimal for measuring 1st change. Six weeks after initiation of chemotherapy with cetuximab for mCRC⁹⁶, 7 weeks after initiation of capecitabine or 5-FU for mCRC⁴² or 8 weeks after initiation of imatinib therapy for GIST⁹⁷ have previously also been used.

In study III, a new lesion may represent an increase from an undetectable sub-cm to 1 cm lesion, and does not necessary indicate more aggressiveness than an increase from 1 cm to 1.2 cm. When a new lesion is reported in a trial, its size is not recorded in the CRFs. Without re-evaluating all imaging, we could not further explore this issue.

In study IV, the data also comes from a relatively restricted number of patients. The cut-off point in case of tumor shrinkage could not be determined because of lack of statistical power. Even though it could be possible, a majority of the patients who developed disease progression received second and later lines of treatments, thus the

relationship between the first tumor shrinkage and OS could be influenced by second and later lines of treatments.

5.4 FUTURE PERSPECTIVES

The findings of this thesis have revealed various limitations in the current tumor response evaluation criteria. Future studies have to be conducted in order to answer the following questions:

1. Is the first change applicable for molecular targeted anticancer agents like for cytotoxic anticancer agents?
 - The first change in metastatic colorectal cancer treated with the Nordic VII study including cetuximab (Erbix[®], anti-EGFR antibody) is planned.
2. Can the first change oriented tumor monitoring be beneficial for cancer care?
 - Enable better follow-up and care and ultimately prolong the overall survival?
 - Omit unnecessary CT or PET-CT examinations and economical burden?
 - Possibility to develop a better phase III clinical trial model based on the first change which providing better planning and minimizing the failure of phase III trials?
3. Does the tumor size really matter or not?
 - Is the change of tumor size before and after treatment really relevant for the prognosis?
 - Do measurements based on the Euclidian geometry reflect the biological behavior of a tumor?
 - Do the fractal dimension rather than Euclidian geometry serve a better indicator to evaluate tumors?

The Euclidian geometry is suitable to measure rather simple structures with smooth surface, e.g. spheres, which never be seen in tumors. Tumors have irregular and heterogeneous structures which can be measured by Fractal geometry rather than Euclidian geometry. Fractal geometry has been used to express the complex network of neurons, blood vessels and tumors⁹⁸⁻¹⁰⁰.

6 CONCLUSIONS

1. WHO-criteria and RECIST contain a large source of random errors which may cause variability deteriorating the reliability of clinical trials.
2. The initial change in tumor size at the first follow-up CT examinations after initiation of treatment correlates with overall survival in mCRC and MBC.
3. The response evaluation must be more focused on detecting appearance of new lesion or the progression of non-target lesion. These indicates particularly poor prognosis. Anatomical imaging such as CT is still useful for response evaluation with wide accessibility. The current process of how to read-out response should be improved.

7 APPENDIX

7.1 ILLUSTRATIVE CHART FOR TUMOR EVALUATION

An illustrative chart for tumor response evaluation according to RECIST and WHO is shown.

Data are from a patient with multiple lung, liver, lymph node and bone metastasis.

Extracted from RadioGraphics 2008¹⁰¹, copyright 2012 by permission of The

Radiological Society of North America

(RSNA[®]).

Baseline Study				Follow-up study		
1 (yy-mm-dd)	Tumor Involvement	Presence of "non-target" lesions	reason	2 (yy-mm-dd)	3 (yy-mm-dd)	4 (yy-mm-dd)
(1) Primary	Y (N) / U	Y (N) / U	<10mm / >5 / other	Appearance of new lesion		
(2) LN	(Y) / N / U	Y (N) / U	<10mm / >5 / other	No ↓ Yes ↓ PD		
(3) Lung	(Y) / N / U	(Y) / N / U	≤10mm >5 other	No ↓ Yes ↓ PD	No ↓ Yes ↓ PD	No ↓ Yes ↓ PD
(4) Liver	(Y) / N / U	(Y) / N / U	≤10mm >5 other	No ↓ Yes ↓ PD	No ↓ Yes ↓ PD	No ↓ Yes ↓ PD
(5) Brain	Y / N (U)	Y / N / U	<10mm / >5 / other	Unequivocal progression of non-target lesion		
(6) Other site	Y (N) / U	Y (N) / U	<10mm / >5 / other	No ↓ Yes ↓ PD		
Bone	(Y) / N / U	Where? <i>Rt. iliac</i>		No ↓ Yes ↓ PD	No ↓ Yes ↓ PD	No ↓ Yes ↓ PD
Pleural effusion	Y (N) / U	Malignancy related? Y / N / U		No ↓ Yes ↓ PD	No ↓ Yes ↓ PD	No ↓ Yes ↓ PD
Ascites	Y (N) / U	Malignancy related? Y / N / U		No ↓ Yes ↓ PD	No ↓ Yes ↓ PD	No ↓ Yes ↓ PD
Other "truly" nontarget lesions	Leptomeningeal / Inflammatory breast disease / Lymphangitis / cystic / necrotic / Lesion in the post irradiated area / others ()			Measure target lesion		
Target	Site *	comment	Size (mm)	Size (mm)	Size (mm)	Size (mm)
1	(1) / (2) / (3) / (4) / (5) / (6)	S5	26.2 (x 21.5) #	24.5 (21.0)	22.7 (21.0)	27.6 (24.1)
2	(1) / (2) / (3) / (4) / (5) / (6)	S4	22.3 (x 17.7)	19.9 (15.3)	16.5 (14.3)	23.1 (20.3)
3	(1) / (2) / (3) / (4) / (5) / (6)	S3	20.9 (x 13.9)	15.9 (10.4)	10.4 (9.4)	22.8 (17.3)
4	(1) / (2) / (3) / (4) / (5) / (6)	Rt. S5 SW	15.9 (x 12.0)	12.4 (9.8)	9.8 (8.4)	19.3 (15.8)
5	(1) / (2) / (3) / (4) / (5) / (6)	paraAo	14.5 (x 13.4)	12.4 (7.4)	12.6 (11.3)	17.3 (14.4)
6	(1) / (2) / (3) / (4) / (5) / (6)	Rt. S10 SW	12.1 (x 11.0)	6.8 (3.2)	5.2 (4.2)	15.6 (12.8)
7	(1) / (2) / (3) / (4) / (5) / (6)					
8	(1) / (2) / (3) / (4) / (5) / (6)					
9	(1) / (2) / (3) / (4) / (5) / (6)					
10	(1) / (2) / (3) / (4) / (5) / (6)					
Sum (WHO)			111.8 (1765.4)	91.8 (1218.3)	77.2 (1057.2)	125.4 (2277.1)
Evaluation (WHO)			Ratio (%)	-18 (-31)	-31 (-41)	+62 (+115)
Comment			Category	SD (SD)	PR (PD)	PD (PD)
Comment			hamartoma in the left lung			

* each number under "Site" corresponds to the above-mentioned number of an organ, e.g., (2)=lymph node, (4)=liver

values in parentheses indicate the longest perpendicular diameter of the target lesion for 2D (WHO) evaluation

Base line evaluation

A: Note the presence or absence of tumor involvement.

B: Start from the largest lesion.

C: Note in which window-setting the lesion is measured.

Follow-up evaluation

D: Before starting measurements, look for new lesions or unequivocal progression of non-target lesions that are critical to follow-up evaluation. The patient will then automatically be classified as PD regardless of other measurements.

E: Even if the target lesions become too small to obtain accurate measurement, it is important to keep on reporting them. *Or use default value of 5 mm (only valid for RECIST 1.1)*

F: In case of disease progression, the smallest sum of the longest diameters since the trial started should be referred rather than the baseline sum of the longest diameters.

G: Since lymph node metastasis often change size along their short axis, measurement of the longest diameter in the trans-axial plane may conceal change of size. *Lymph nodes should be measured in the short axis if RECIST 1.1 is used in the trial.*

In WHO-criteria, if the product of longest diameter and longest perpendicular diameter increased more than 25% in any lesion, the patient should be categorized as PD. On the other hand, in RECIST, progression in any lesion without an overall 20% increase in the sum of longest diameter is not considered progression.

As in this patient shown above, the lymph node metastasis increases more than 25% at the third follow-up. The patient should be classified as PD according to WHO-criteria. However, in RECIST, the sum of the longest diameters is more than 30% reduction of baseline sum and the patient is classified as PR.

8 ACKNOWLEDGEMENTS

I wish to express my sincere gratitude for the following persons whose help has been indispensable. Without you, I am sure, this would never come true.

My principal supervisor Lennart Blomqvist, who keeps on encouraging me and protecting me. Always paving the “Royal Road” for PhD candidates like me and guiding not to stray from the path. I cannot find any word to describe how I thank you! I have learnt a lot.

My co-supervisor Professor Hans Jacobsson, a brilliant person with abundant knowledge in many subjects and great sense of humor. The whole idea about this tumor response evaluation came from you. Without your support, I could not survive here.

My co-supervisor Hirofumi Fujii, one of the smartest people I have ever met. I am really lucky to have you as my supervisor ever since I was in Keio University in Japan.

Professor Bengt Glimelius, you are so brilliant, needless to explain! Thank you very much for your kindly understanding to my primitive ideas of the first change and thoughtful comments, without them I were still be in the depths. I am looking forward to our future collaborations

Professor Anders Sundin, a great educator as well as a brilliant scientist. Always keep on encouraging me and keep on trusting me. I will never forget your heartwarming words.

Thomas Hatschek, a great clinician and a scientist who keep on encouraging me from the beginning when I just came to Sweden. One of the key person who let me in this academic world. I am really lucky to meet you.

Maria Liljefors, my external mentor and one of my honorable role-model as a scientist, a woman and a mother. You rescued me, comforted me and encouraged me when I was in a terrible deep ditch.

Elisabeth Berg, the greatest statistician of all projects, I also do appreciate your patience to try to understand what I say in my bad Swedish or even worse in English.

Michael Torkzad, a distinguished colleague and a friend, I am looking forward to future collaborations.

Ann-Britt Wikstrom, excellent administrative of IMMK, always supporting me and guiding me when I was nearly drowning in a bunch of complex and long paper works.

Clementine Molin, always elegant, very smart and keep on encouraging me and telling me the importance of these projects for the development of cancer drugs.

The present head of Department of Diagnostic Radiology, Henry Lindholm and the former head of Department of Diagnostic Radiology, Lott Bergstrand for your kind arrangements and support.

Boel Söderén, IngMarie Meyer, Katarina Bodén, you are my sisters in Sweden!

Without your friendship I had been running back to Japan. Your jokes, a lot of laughs and encouragements keep me anchored here.

Veli Söderlund, Jan Bohlin, HaakonViland, Magnus Tengvar, Erik Rollvén, Mikael Skorpil wonderful colleagues and friends. What a joy to work together with you!

Roberto, Yvonne, Manijeh, Mia and all fantastic MR staffs, thanks making me always happy!

Toya, Kirsi, Solweig for your great hospitality, it is always great pleasure to talk with you!

Noriko Kamata who is my "master Yoda" not only in my medical carrier in the radiology but in my life. Without you, I never be like me at present, yes, this is all your presents for me, thank you!

Fumikazu Sakai, Takashi Ushimi, both great radiologists and scientists. I appreciate your abundant knowledge that makes many radiology congresses fruitful.

All my friends both in Japan and in Sweden and all colleagues and staffs at the Department of Diagnostic Radiology, Karolinska University Hospital Solna and Department of Diagnostic Radiology Tokyo Metropolitan Cancer and Infectious Disease Center Komagome Hospital, Tokyo, Japan.

My father, Masayoshi Suzuki, he passed away in 2010. I am so sorry that I could not show you my thesis.

My mother, Tomoko Suzuki, who always pushes my back to move forward, always accept whatever I decide. You are the greatest mother and grandmother in the world.

And last but not least my beloved daughter, Sakiko. Saki had no idea how long she would stay in Sweden or how she would be like a "cosmopolitan" when she was taken here at the age of 5 by idiotic mother. Thank you for creating the cover jacket of this thesis. What a great privilege to be your mother!

9 REFERENCES

1. Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. *J Clin Oncol*. Nov 1983;1(11):710-719.
2. Moertel CG, Hanley JA. The effect of measuring error on the results of therapeutic trials in advanced cancer. *Cancer*. Jul 1976;38(1):388-394.
3. *WHO handbook for reporting results of cancer treatment*. Vol Available from: <http://whqlibdoc.who.int/publications/9241700483.pdf>. Geneva: World Health Organization; 1979.
4. World Health Organization. *WHO handbook for reporting results of cancer treatment*. Geneva Albany, N.Y.: World Health Organization ; sold by WHO Publications Centre USA; 1979.
5. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst*. Feb 2 2000;92(3):205-216.
6. Nygren P, Blomqvist L, Bergh J, Astrom G. Radiological assessment of tumour response to anti-cancer drugs: time to reappraise. *Acta Oncol*. 2008;47(2):316-318.
7. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. Jan 2009;45(2):228-247.
8. James K, Eisenhauer E, Christian M, Terenziani M, Vena D, Muldal A, et al. Measuring response in solid tumors: unidimensional versus bidimensional measurement. *J Natl Cancer Inst*. Mar 17 1999;91(6):523-528.
9. Park JO, Lee SI, Song SY, Kim K, Kim WS, Jung CW, et al. Measuring response in solid tumors: comparison of RECIST and WHO response criteria. *Jpn J Clin Oncol*. Oct 2003;33(10):533-537.
10. Trillet-Lenoir V, Freyer G, Kaemmerlen P, Fond A, Pellet O, Lombard-Bohas C, et al. Assessment of tumour response to chemotherapy for metastatic colorectal cancer: accuracy of the RECIST criteria. *Br J Radiol*. Nov 2002;75(899):903-908.
11. Fabel M, von Tengg-Kobligk H, Giesel FL, Bornemann L, Dicken V, Kopp-Schneider A, et al. Semi-automated volumetric analysis of lymph node metastases in patients with malignant melanoma stage III/IV-A feasibility study. *Eur Radiol*. Jun 2008;18(6):1114-1122.
12. Sohaib SA, Turner B, Hanson JA, Farquharson M, Oliver RT, Reznick RH. CT assessment of tumour response to treatment: comparison of linear, cross-sectional and volumetric measures of tumour size. *Br J Radiol*. Nov 2000;73(875):1178-1184.
13. Tran LN, Brown MS, Goldin JG, Yan X, Pais RC, McNitt-Gray MF, et al. Comparison of treatment response classifications between unidimensional, bidimensional, and volumetric measurements of metastatic lung lesions on chest computed tomography. *Acad Radiol*. Dec 2004;11(12):1355-1360.
14. Marten K, Auer F, Schmidt S, Kohl G, Rummeny EJ, Engelke C. Inadequacy of manual measurements compared to automated CT volumetry in assessment of treatment response of pulmonary metastases using RECIST criteria. *Eur Radiol*. Apr 2006;16(4):781-790.
15. Jacene HA, Lebouilleux S, Baba S, Chatzifotiadis D, Goudarzi B, Teytelbaum O, et al. Assessment of interobserver reproducibility in quantitative 18F-FDG PET and CT measurements of tumor response to therapy. *J Nucl Med*. Nov 2009;50(11):1760-1769.
16. Suzuki C, Blomqvist L, Sundin A, Jacobsson H, Bystrom P, Berglund A, et al. The initial change in tumor size predicts response and survival in patients with

- metastatic colorectal cancer treated with combination chemotherapy. *Ann Oncol.* Apr 2012;23(4):948-954.
17. Zacharia TT, Saini S, Halpern EF, Sumner JE. CT of colon cancer metastases to the liver using modified RECIST criteria: determining the ideal number of target lesions to measure. *AJR Am J Roentgenol.* Apr 2006;186(4):1067-1070.
 18. Schwartz LH, Mazumdar M, Brown W, Smith A, Panicek DM. Variability in response assessment in solid tumors: effect of number of lesions chosen for measurement. *Clin Cancer Res.* Oct 1 2003;9(12):4318-4323.
 19. Graf W, Pahlman L, Bergstrom R, Glimelius B. The relationship between an objective response to chemotherapy and survival in advanced colorectal cancer. *Br J Cancer.* Sep 1994;70(3):559-563.
 20. Ratain MJ, Eckhardt SG. Phase II studies of modern drugs directed against new targets: if you are fazed, too, then resist RECIST. *J Clin Oncol.* Nov 15 2004;22(22):4442-4445.
 21. Buyse M, Thirion P, Carlson RW, Burzykowski T, Molenberghs G, Piedbois P. Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: a meta-analysis. Meta-Analysis Group in Cancer. *Lancet.* Jul 29 2000;356(9227):373-378.
 22. De Roock W, Piessevaux H, De Schutter J, Janssens M, De Hertogh G, Personeni N, et al. KRAS wild-type state predicts survival and is associated to early radiological response in metastatic colorectal cancer treated with cetuximab. *Ann Oncol.* Mar 2008;19(3):508-515.
 23. Burzykowski T, Buyse M, Piccart-Gebhart MJ, Sledge G, Carmichael J, Luck HJ, et al. Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer. *J Clin Oncol.* Apr 20 2008;26(12):1987-1992.
 24. Brufsky AM, Hurvitz S, Perez E, Swamy R, Valero V, O'Neill V, et al. RIBBON-2: a randomized, double-blind, placebo-controlled, phase III trial evaluating the efficacy and safety of bevacizumab in combination with chemotherapy for second-line treatment of human epidermal growth factor receptor 2-negative metastatic breast cancer. *J Clin Oncol.* Nov 10 2011;29(32):4286-4293.
 25. Dimitrakopoulou-Strauss A, Strauss LG, Burger C, Ruhl A, Irmgartinger G, Stremmel W, et al. Prognostic aspects of 18F-FDG PET kinetics in patients with metastatic colorectal carcinoma receiving FOLFOX chemotherapy. *J Nucl Med.* Sep 2004;45(9):1480-1487.
 26. de Geus-Oei LF, van Laarhoven HW, Visser EP, Hermsen R, van Hoorn BA, Kamm YJ, et al. Chemotherapy response evaluation with FDG-PET in patients with colorectal cancer. *Ann Oncol.* Feb 2008;19(2):348-352.
 27. Bystrom P, Berglund A, Garske U, Jacobsson H, Sundin A, Nygren P, et al. Early prediction of response to first-line chemotherapy by sequential [18F]-2-fluoro-2-deoxy-D-glucose positron emission tomography in patients with advanced colorectal cancer. *Ann Oncol.* Jun 2009;20(6):1057-1061.
 28. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer.* Dec 15 2010;127(12):2893-2917.
 29. Boyle P, Levin B, International Agency for Research on Cancer., World Health Organization. *World cancer report 2008.* Lyon Geneva: International Agency for Research on Cancer ; Distributed by WHO Press; 2008.
 30. Farmer P, Frenk J, Knaul FM, Shulman LN, Alleyne G, Armstrong L, et al. Expansion of cancer care and control in countries of low and middle income: a call to action. *Lancet.* Oct 2 2010;376(9747):1186-1193.
 31. Smith BD, Smith GL, Hurria A, Hortobagyi GN, Buchholz TA. Future of cancer incidence in the United States: burdens upon an aging, changing nation. *J Clin Oncol.* Jun 10 2009;27(17):2758-2765.
 32. Hricak H. Oncologic imaging: a guiding hand of personalized cancer care. *Radiology.* Jun 2011;259(3):633-640.
 33. Forner A, Ayuso C, Varela M, Rimola J, Hessheimer AJ, de Lope CR, et al. Evaluation of tumor response after locoregional therapies in hepatocellular

- carcinoma: are response evaluation criteria in solid tumors reliable? *Cancer*. Feb 1 2009;115(3):616-623.
34. Schrag D. The price tag on progress--chemotherapy for colorectal cancer. *N Engl J Med*. Jul 22 2004;351(4):317-319.
 35. DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. *J Health Econ*. Mar 2003;22(2):151-185.
 36. Jefford M, Savulescu J, Thomson J, Schofield P, Mileskin L, Agalinos E, et al. Medical paternalism and expensive unsubsidised drugs. *BMJ*. Nov 5 2005;331(7524):1075-1077.
 37. Johnson JR, Williams G, Pazdur R. End points and United States Food and Drug Administration approval of oncology drugs. *J Clin Oncol*. Apr 1 2003;21(7):1404-1411.
 38. Dagher R, Johnson J, Williams G, Keegan P, Pazdur R. Accelerated approval of oncology products: a decade of experience. *J Natl Cancer Inst*. Oct 20 2004;96(20):1500-1509.
 39. Verma S, McLeod D, Batist G, Robidoux A, Martins IR, Mackey JR. In the end what matters most? A review of clinical endpoints in advanced breast cancer. *Oncologist*. 2011;16(1):25-35.
 40. Karrison TG, Maitland ML, Stadler WM, Ratain MJ. Design of phase II cancer trials using a continuous endpoint of change in tumor size: application to a study of sorafenib and erlotinib in non small-cell lung cancer. *J Natl Cancer Inst*. Oct 3 2007;99(19):1455-1461.
 41. Ratain MJ. Phase II oncology trials: let's be positive. *Clin Cancer Res*. Aug 15 2005;11(16):5661-5662.
 42. Claret L, Girard P, Hoff PM, Van Cutsem E, Zuideveld KP, Jorga K, et al. Model-based prediction of phase III overall survival in colorectal cancer on the basis of phase II tumor dynamics. *J Clin Oncol*. 2009;27(25):4103-4108. Epub 2009 Jul 4127.
 43. Goffin J, Baral S, Tu D, Nomikos D, Seymour L. Objective responses in patients with malignant melanoma or renal cell cancer in early clinical studies do not predict regulatory approval. *Clin Cancer Res*. Aug 15 2005;11(16):5928-5934.
 44. Richter WS. Imaging biomarkers as surrogate endpoints for drug development. *Eur J Nucl Med Mol Imaging*. Jul 2006;33(Supplement 13):6-10.
 45. Buyse M, Sargent DJ, Grothey A, Matheson A, de Gramont A. Biomarkers and surrogate end points--the challenge of statistical validation. *Nat Rev Clin Oncol*. Jun 2010;7(6):309-317.
 46. Amado RG, Wolf M, Peeters M, Van Cutsem E, Siena S, Freeman DJ, et al. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J Clin Oncol*. Apr 1 2008;26(10):1626-1634.
 47. Rastelli F, Crispino S. Factors predictive of response to hormone therapy in breast cancer. *Tumori*. May-Jun 2008;94(3):370-383.
 48. McKee AE, Farrell AT, Pazdur R, Woodcock J. The role of the U.S. Food and Drug Administration review process: clinical trial endpoints in oncology. *Oncologist*. 2010;15 Suppl 1:13-18.
 49. Pazdur R. Endpoints for assessing drug activity in clinical trials. *Oncologist*. 2008;13 Suppl 2:19-21.
 50. Saad ED, Katz A. Progression-free survival and time to progression as primary end points in advanced breast cancer: often used, sometimes loosely defined. *Ann Oncol*. Mar 2009;20(3):460-464.
 51. Saad ED, Katz A, Hoff PM, Buyse M. Progression-free survival as surrogate and as true end point: insights from the breast and colorectal cancer literature. *Ann Oncol*. Jan 2010;21(1):7-12.
 52. Gilman A, Philips FS. The Biological Actions and Therapeutic Applications of the B-Chloroethyl Amines and Sulfides. *Science*. Apr 5 1946;103(2675):409-436.
 53. Jacobson LO, Spurr CL, et al. Nitrogen mustard therapy; studies on the effect of methyl-bis (beta-chloroethyl) amine hydrochloride on neoplastic diseases and allied disorders of the hemopoietic system. *J Am Med Assoc*. Oct 5 1946;132:263-271.

54. Warr D, McKinney S, Tannock I. Influence of measurement error on assessment of response to anticancer chemotherapy: proposal for new criteria of tumor response. *J Clin Oncol*. Sep 1984;2(9):1040-1046.
55. Thiesse P, Ollivier L, Di Stefano-Louineau D, Negrier S, Savary J, Pignard K, et al. Response rate accuracy in oncology trials: reasons for interobserver variability. Groupe Francais d'Immunotherapie of the Federation Nationale des Centres de Lutte Contre le Cancer. *J Clin Oncol*. Dec 1997;15(12):3507-3514.
56. Lavin PT, Flowerdew G. Studies in variation associated with the measurement of solid tumors. *Cancer*. Sep 1 1980;46(5):1286-1290.
57. James K, Eisenhauer E, Therasse P. Re: Measure once or twice--does it really matter? *J Natl Cancer Inst*. Oct 20 1999;91(20):1780-1781.
58. Prasad SR, Jhaveri KS, Saini S, Hahn PF, Halpern EF, Sumner JE. CT tumor measurement for therapeutic response assessment: comparison of unidimensional, bidimensional, and volumetric techniques initial observations. *Radiology*. Nov 2002;225(2):416-419.
59. Mazumdar M, Smith A, Schwartz LH. A statistical simulation study finds discordance between WHO criteria and RECIST guideline. *J Clin Epidemiol*. Apr 2004;57(4):358-365.
60. Warren KE, Patronas N, Aikin AA, Albert PS, Balis FM. Comparison of one-, two-, and three-dimensional measurements of childhood brain tumors. *J Natl Cancer Inst*. Sep 19 2001;93(18):1401-1405.
61. Hilsenbeck SG, Von Hoff DD. Measure once or twice -- does it really matter? *J Natl Cancer Inst*. Mar 17 1999;91(6):494-495.
62. Padhani AR, Ollivier L. The RECIST (Response Evaluation Criteria in Solid Tumors) criteria: implications for diagnostic radiologists. *Br J Radiol*. Nov 2001;74(887):983-986.
63. Erasmus JJ, Gladish GW, Broemeling L, Sabloff BS, Truong MT, Herbst RS, et al. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. *J Clin Oncol*. Jul 1 2003;21(13):2574-2582.
64. Oxnard GR, Zhao B, Sima CS, Ginsberg MS, James LP, Lefkowitz RA, et al. Variability of lung tumor measurements on repeat computed tomography scans taken within 15 minutes. *J Clin Oncol*. Aug 10 2011;29(23):3114-3119.
65. Mazumdar M, Smith A, Debroy PP, Schwartz LH. A theoretical approach to choosing the minimum number of multiple tumors required for assessing treatment response. *J Clin Epidemiol*. Feb 2005;58(2):150-153.
66. Seddon BM, Workman P. The role of functional and molecular imaging in cancer drug discovery and development. *Br J Radiol*. 2003;76 Spec No 2:S128-138.
67. Moskowitz CS, Jia X, Schwartz LH, Gonen M. A simulation study to evaluate the impact of the number of lesions measured on response assessment. *Eur J Cancer*. Jan 2009;45(2):300-310.
68. Schwartz LH, Bogaerts J, Ford R, Shankar L, Therasse P, Gwyther S, et al. Evaluation of lymph nodes with RECIST 1.1. *Eur J Cancer*. Jan 2009;45(2):261-267.
69. Lavin PT. An alternative model for the evaluation of antitumor activity. *Cancer Clin Trials*. Winter 1981;4(4):451-457.
70. Coleman RE, Laymon CM, Turkington TG. FDG imaging of lung nodules: a phantom study comparing SPECT, camera-based PET, and dedicated PET. *Radiology*. Mar 1999;210(3):823-828.
71. Su H, Bodenstein C, Dumont RA, Seimbille Y, Dubinett S, Phelps ME, et al. Monitoring tumor glucose utilization by positron emission tomography for the prediction of treatment response to epidermal growth factor receptor kinase inhibitors. *Clin Cancer Res*. Oct 1 2006;12(19):5659-5667.
72. Weber WA, Figlin R. Monitoring Cancer Treatment with PET/CT: Does It Make a Difference? *J Nucl Med*. Jan 2007;48 Suppl 1:36S-44S.
73. Delbeke D. Oncological applications of FDG PET imaging. *J Nucl Med*. Oct 1999;40(10):1706-1715.
74. Young H, Baum R, Cremerius U, Herholz K, Hoekstra O, Lammertsma AA, et al. Measurement of clinical and subclinical tumour response using [18F]-

- fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. *Eur J Cancer*. Dec 1999;35(13):1773-1782.
75. Shankar LK, Hoffman JM, Bacharach S, Graham MM, Karp J, Lammertsma AA, et al. Consensus recommendations for the use of 18F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute Trials. *J Nucl Med*. Jun 2006;47(6):1059-1066.
 76. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *J Nucl Med*. May 2009;50 Suppl 1:122S-150S.
 77. Shepherd FA, Rodrigues Pereira J, Ciuleanu T, Tan EH, Hirsh V, Thongprasert S, et al. Erlotinib in previously treated non-small-cell lung cancer. *N Engl J Med*. Jul 14 2005;353(2):123-132.
 78. Thatcher N, Chang A, Parikh P, Rodrigues Pereira J, Ciuleanu T, von Pawel J, et al. Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomised, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer). *Lancet*. Oct 29-Nov 4 2005;366(9496):1527-1537.
 79. Koh DM, Scurr E, Collins D, Kanber B, Norman A, Leach MO, et al. Predicting response of colorectal hepatic metastasis: value of pretreatment apparent diffusion coefficients. *AJR Am J Roentgenol*. Apr 2007;188(4):1001-1008.
 80. Koh DM, Takahara T, Imai Y, Collins DJ. Practical aspects of assessing tumors using clinical diffusion-weighted imaging in the body. *Magn Reson Med Sci*. 2007;6(4):211-224.
 81. Suzuki C, Torkzad MR, Jacobsson H, Astrom G, Sundin A, Hatschek T, et al. Interobserver and intraobserver variability in the response evaluation of cancer therapy according to RECIST and WHO-criteria. *Acta Oncol*. May 2010;49(4):509-514.
 82. Kundel HL, Polansky M. Measurement of observer agreement. *Radiology*. Aug 2003;228(2):303-308.
 83. Buyse M, Piedbois P. On the relationship between response to treatment and survival time. *Stat Med*. Dec 30 1996;15(24):2797-2812.
 84. Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response and other comparisons of time-to-event by outcome variables. *J Clin Oncol*. Aug 20 2008;26(24):3913-3915.
 85. Campbell MJ, Walters SJ, Machin D. *Medical statistics : a textbook for the health sciences*. 4th ed. Chichester, England ; Hoboken, NJ: Wiley; 2007.
 86. Glimelius B, Sorbye H, Balteskard L, Bystrom P, Pfeiffer P, Tveit K, et al. A randomized phase III multicenter trial comparing irinotecan in combination with the Nordic bolus 5-FU and folinic acid schedule or the bolus/infused de Gramont schedule (Lv5FU2) in patients with metastatic colorectal cancer. *Ann Oncol*. May 2008;19(5):909-914.
 87. Hatschek T, Carlsson L, Einbeigi Z, Lidbrink E, Linderholm B, Lindh B, et al. Individually tailored treatment with epirubicin and paclitaxel with or without capecitabine as first-line chemotherapy in metastatic breast cancer: a randomized multicenter trial. *Breast Cancer Res Treat*. Feb 2012;131(3):939-947.
 88. Darkeh MH, Suzuki C, Torkzad MR. The minimum number of target lesions that need to be measured to be representative of the total number of target lesions (according to RECIST). *Br J Radiol*. Aug 2009;82(980):681-686.
 89. Skougard K, McCullagh MJ, Nielsen D, Hendel HW, Jensen BV, Johannesen HH. Observer variability in a phase II trial - assessing consistency in RECIST application. *Acta Oncol*. Mar 21 2012.
 90. Ford R, Schwartz L, Dancy J, Dodd LE, Eisenhauer EA, Gwyther S, et al. Lessons learned from independent central review. *Eur J Cancer*. Jan 2009;45(2):268-274.
 91. Marten K, Auer F, Schmidt S, Rummeny EJ, Engelke C. Automated CT volumetry of pulmonary metastases: the effect of a reduced growth threshold

- and target lesion number on the reliability of therapy response assessment using RECIST criteria. *Eur Radiol.* Oct 2007;17(10):2561-2571.
92. Honda O, Sumikawa H, Johkoh T, Tomiyama N, Mihara N, Inoue A, et al. Computer-assisted lung nodule volumetry from multi-detector row CT: influence of image reconstruction parameters. *Eur J Radiol.* Apr 2007;62(1):106-113.
 93. Zhao B, Schwartz LH, Moskowitz CS, Wang L, Ginsberg MS, Cooper CA, et al. Pulmonary metastases: effect of CT section thickness on measurement--initial experience. *Radiology.* Mar 2005;234(3):934-939.
 94. Schwartz LH, Ginsberg MS, DeCorato D, Rothenberg LN, Einstein S, Kijewski P, et al. Evaluation of tumor measurements in oncology: use of film-based and electronic techniques. *J Clin Oncol.* May 2000;18(10):2179-2184.
 95. Richey EA, Lyons EA, Nebeker JR, Shankaran V, McKoy JM, Luu TH, et al. Accelerated approval of cancer drugs: improved access to therapeutic breakthroughs or early release of unsafe and ineffective drugs? *J Clin Oncol.* Sep 10 2009;27(26):4398-4405.
 96. Piessevaux H, Buyse M, De Roock W, Prenen H, Schlichting M, Van Cutsem E, et al. Radiological tumor size decrease at week 6 is a potent predictor of outcome in chemorefractory metastatic colorectal cancer treated with cetuximab (BOND trial). *Ann Oncol.* Aug 2009;20(8):1375-1382.
 97. Choi H, Charnsangavej C, Faria SC, Macapinlac HA, Burgess MA, Patel SR, et al. Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. *J Clin Oncol.* May 1 2007;25(13):1753-1759.
 98. Mandelbrot B. How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science.* May 5 1967;156(3775):636-638.
 99. Mandelbrot BB. Stochastic models for the Earth's relief, the shape and the fractal dimension of the coastlines, and the number-area rule for islands. *Proc Natl Acad Sci U S A.* Oct 1975;72(10):3825-3828.
 100. Cross SS. Fractals in pathology. *J Pathol.* May 1997;182(1):1-8.
 101. Suzuki C, Jacobsson H, Hatschek T, Torkzad MR, Boden K, Eriksson-Alm Y, et al. Radiologic measurements of tumor response to treatment: practical approaches and limitations. *Radiographics.* Mar-Apr 2008;28(2):329-344.